
WATER

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm
on: Fri Sep 17 21:10:05 1999; MasPar time 39.18 Seconds
Tabular output not generated. 1020.796 Million cell updates/sec

Title: >US-09-041-994-2
Description: (1-1415) from US09041994.pgp
Perfect Score: 9849
Sequence: 1 MSGLGENLDPLASDSRKRRL.....MNMNPMXGMPGPDQKYC 1415

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
l:swissprot

Statistics: Mean 56.586; Variance 121.726; scale 0.465

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	232	2.4	805	1	AHR_MOUSE	9.74e-17
2	232	2.4	848	1	AHR_HUMAN	9.74e-17
3	239	2.4	853	1	AHR_RAT	7.91e-18
4	225	2.3	339	1	TF2D_HUMAN	1.17e-15
5	224	2.3	642	1	ARNT_DROME	1.67e-15
6	230	2.3	644	1	BTB_DROME	1.99e-16
7	227	2.3	758	1	Y438_YEAST	5.77e-16
8	226	2.3	905	1	SNF5_YEAST	8.23e-16
9	229	2.3	1794	1	YAV1_SCHPO	2.84e-16
10	219	2.2	262	1	GBA1_WHEAT	9.70e-15
11	215	2.2	313	1	GBA7_WHEAT	3.93e-14
12	221	2.2	609	1	OPA_DROME	4.81e-15
13	213	2.2	2124	1	Y192_HUMAN	7.89e-14
14	217	2.2	2175	1	HMCU_DROME	1.96e-14
15	210	2.1	360	1	MDL1_HUMAN	2.23e-13
16	205	2.1	467	1	INVO_MOUSE	1.25e-12
17	209	2.1	594	1	NPAL_MOUSE	3.15e-13
18	205	2.1	648	1	KAPC_DICDI	1.25e-12
19	204	2.1	708	1	GBF_DICDI	1.76e-12
20	202	2.1	870	1	PAS1_HUMAN	3.49e-12
21	211	2.1	874	1	PAS1_MOUSE	1.58e-12
22	204	2.1	966	1	SSN6_YEAST	1.76e-12
23	204	2.1	1081	1	GALY_YEAST	1.76e-12

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL: D38417; G1215804; -
DR EMBL: M94623; G192101; -
DR EMBL: D38416; -; NOT_ANNOTATED_CDS.
DR PIR: JQ1485; JQ1485.
DR PIR: A46266; A46266.
DR MGD; MGI:105043; AHR.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
DR PFAM; PF00785; PAC; 1.
DR PFAM; PF00989; PAC; 1.
DR TRANSFAC; T00018; -
DR TRANSFAC; T00194; -
KW TRANSCRIPTION REGULATION; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.
FT PROPEP 1 9
CHAIN 10 805
DNA_BIND 27 39
BASIC DOMAIN
DOMAIN 40 80
HELI-LOOP-HELIX MOTIF (BY SIMILARITY).
FT REPEAT 111 177
PAS-1.
FT REPEAT 269 336
PAS-2.
FT DOMAIN 342 380
PAC MOTIF.
FT DOMAIN 594 648
GLN-RICH.
FT VARIANT 324 324
I -> M (IN DBA/2J).
FT VARIANT 351 351
H -> N (IN DBA/2J).
FT VARIANT 375 375
A -> V (IN DBA/2J).
FT VARIANT 471 471
L -> P (IN DBA/2J).
FT VARIANT 533 533
S -> P (IN DBA/2J).
FT VARIANT 589 589
L -> M (IN DBA/2J).
FT VARIANT 805 805
S -> S (IN DBA/2J).
FT CONFLICT 74 74
S -> T (IN REF. 2).
SQ SEQUENCE 805 AA; 90337 MW; 1998DIAA CRC32;

Query Match 2.4%; Score 232; DB 1; Length 805;
Best Local Similarity 30.3%; Pred. No. 9.74e-17;
Matches 46; Conservative 46; Mismatches 52; Indels 8; Gaps 6;

Db 26 PARGIKSNPSKRRDRRLNTELDRLASLLP--FPQ-DVIN-KLDKLSVLRSLVSLRAKSF 81

QY 25 PGQGLTCSGKRRRQESKYIEELAEISANLSDIDNFNVKPKCAILKETVRIQRIKE 84

Db 82 FDVALKSTPADRNGGQDCQACQ--IRWDQLQEGEFLQLQALNGFLVVTADALVFFYASSTI 140

QY 85 QGKTI-SNDDDVOKADVSTGQGVDRKSL--GPLLLQALDGLFVFNREANIVFVSENV 141

QY 141 QDYLGFSQSDVIHQSVYELIHTEDRAEFQRL 172

QY 142 TQYLYQKQEDLVNTSVYNILHEEDRKDFLKNL 173

RESULT 2
ID AHR_HUMAN STANDARD; PRT; 848 AA.

AC P35869;

DT 01-JUN-1994 (REL. 29, CREATED)

DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE AHR RECEPTOR (ARYL HYDROCARBON RECEPTOR).

GN AHR.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

CC [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;

RX MEDLINE; 93347997.

RA ITOH S., KAWATAKI T.;

RT "Human Ah receptor cDNA: analysis for highly conserved sequences.";

RL NUCLEIC ACIDS RES. 21:3578-3578(1993).

RP SEQUENCE FROM N.A.

RX MEDLINE; 94067047.
RA DOLWICK K.M., SCHMIDT J.V., CARVER L.A., SWANSON H.I., BRADFIELD C.A.;
RT "Cloning and expression of a human Ah receptor cDNA.";
RL MOL. PHARMACOL. 44:911-917(1993).
RN [3]

RP SEQUENCE FROM N.A.

RA ANTONIOU B., SMITH A., ELLIOTT G., KRAMER J.;

RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -!- FUNCTION: THE AH RECEPTOR ACTIVATED IN ASSOCIATION WITH A LIGAND

CC IS CONSIDERED TO TRANSLOCATE FROM CYTOPLASM TO NUCLEUS AND

CC ENHANCE THE TRANSCRIPTION OF THE GENES BY BINDING TO THE XRE

CC SEQUENCE IN THEIR PROMOTER REGIONS. IT MEDIATES BIOCHEMICAL AND

CC TOXIC EFFECTS OF HALOGENATED AROMATIC HYDROCARBONS.

CC -!- SUBCELLULAR LOCATION: FIRST CYTOPLASMIC. UPON BINDING WITH LIGAND

CC AND INTERACTION WITH A HSP90, IT TRANSLOCATES TO THE NUCLEUS.

CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH AN OTHER

CC BHLH PROTEIN. IN THE NUCLEUS, HETERODIMER OF AHR AND ARNT.

CC -!- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.

CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF

CC TRANSCRIPTION FACTORS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D16354; G533324; -

DR EMBL; L19872; G416142; -

DR EMBL; AC003075; G2588638; -

DR MIM; 600253; -

DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.

DR PFAM; PF00785; PAC; 1.

DR PFAM; PF00989; PAC; 1.

DR TRANSFAC; T01795; -

KW TRANSCRIPTION REGULATION; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.

FT DNA_BIND 28 40

FT DOMAIN 41 81

FT REPEAT 113 179

FT REPEAT 275 342

FT DOMAIN 348 386

FT DOMAIN 600 640

FT CONFLICT 807 848

FT L -> FK (IN REF. 1).

SQ SEQUENCE 848 AA; 96147 MW; 51BB7808 CRC32;

Query Match 2.4%; Score 232; DB 1; Length 848;
Best Local Similarity 27.5%; Pred. No. 9.74e-17;
Matches 42; Conservative 52; Mismatches 50; Indels 9; Gaps 7;

Db 27 PARGIKSNPSKRRDRRLNTELDRLASLLP--FPQ-DVIN-KLDKLSVLRSLVSLRAKSF 82

QY 25 PGQGLTCSGKRRRQESKYIEELAEISANLSDIDNFNVKPKCAILKETVRIQRIK- 83

Db 83 FDVALKSSPTERNQDNCRAANFREGNLQE--GEFLQALNGFLVVTADALVFFYASST 141

QY 84 -EGG-KTISNDDDVOKADVSTG-QGVDRKSLGPLLLQALDGLFVFNREANIVFVSENV 140

Db 142 IDYLGFSQSDVIHQSVYELIHTEDRAEFQRL 174

QY 141 TQYLYQKQEDLVNTSVYNILHEEDRKDFLKNL 173

RESULT 3
ID AHR_RAT STANDARD; PRT; 853 AA.

AC P41738;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE AHR RECEPTOR (ARYL HYDROCARBON RECEPTOR).

GN AHR.

QY	85	QGKTI-SNDDVQKADVSSTGGQVDKDSL--GPLLLQALDQFLFVVVNRANIVFVSENV	141
Db	141	QDYLGGQQSDVIHQSVYELIHTEDRAEFQRL	172
QY	142	TOYLQYKQEDLVNTSVYNILHEEDRKDFLKNL	173

RESULT	4		
ID	TF2D_HUMAN	STANDARD;	PRT; 339 AA.
AC	P20226;		
DT	01-FEB-1991	(REL. 17, CREATED)	
DT	01-FEB-1996	(REL. 33, LAST SEQUENCE UPDATE)	
DT	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)	
DT	DE	TRANSCRIPTION INITIATION FACTOR TF2ID (TATA-BOX FACTOR) (TATA	
DE	DE	SEQUENCE-BINDING PROTEIN) (TBP).	
DE	TBP OR TF2ID OR TF2D.		
OS	HOMO SAPIENS (HUMAN).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;		
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE; 90302010.		
RX	KAO C.C., LIEBERMAN P.M., SCHMIDT M.C., ZHOU Q., PEI R., BERK A.J.;		
RA	"Cloning of a transcriptionally active human TATA binding factor.";		
RL	SCIENCE 248:1646-1649(1990).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE; 90326195.		
RX	HOFFMANN A., SINN E., YAMAMOTO T., WANG J., ROY A., HORIKOSHI M.,		
RA	ROEDER R.G.;		
RL	"Highly conserved core domain and unique N terminus with presumptive		
RT	regulatory motifs in a human TATA factor (TF2ID).";		
RL	NATURE 346:387-390(1990).		
RL	[3]		
RN	DOMAINS.		
RX	MEDLINE; 90302006.		
RA	PETERSON M.G., TANESE N., PUGH B.F., TJIAN R.;		
RL	"Functional domains and upstream activation properties of cloned		
RT	human TATA binding protein.";		
RL	SCIENCE 248:1625-1630(1990).		
RL	[4]		
RN	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 159-337.		
RX	MEDLINE; 96209823.		
RX	NIKOLOV D.B., CHEN H., HALAY E.D., HOFFMANN A., ROEDER R.G.,		
RA	BURLEY S.K.;		
RL	"Crystal structure of a human TATA box-binding protein/TATA element		
RT	complex.";		
RL	PROC. NATL. ACAD. SCI. U.S.A. 93:4862-4867(1996).		
RL	[5]		
RN	X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 159-339.		
RX	MEDLINE; 96346176.		
RX	JUO Z.S., CHIU T.K., LETBERMAN P.M., BAIKALOV I., BERK A.J.,		
RA	DICKERSON R.E.;		
RL	"How proteins recognize the TATA box.";		
RT	J. MOL. BIOL. 261:239-254(1996).		
CC	!- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION		
CC	OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TF2ID BINDS		
CC	SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO		
CC	THE POSITION OF TRANSCRIPTION INITIATION.		
CC	!- SUBUNIT: BINDS DNA AS A MONOMER.		
CC	!- SUBCELLULAR LOCATION: NUCLEAR.		
CC	!- SIMILARITY: THE C-TERMINAL 180 RESIDUES ARE EXTREMELY WELL		
CC	CONSERVED IN ALL EUKARYOTIC TF2ID.		
CC	!- SIMILARITY: WEAK, WITH BACTERIAL POLYMERASE SIGMA-FACTORS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		

```

Query Match      2.4%; Score 339; DB 1; Length 853;
Best Local Similarity 30.9%; Pred. No. 7.9ie-18;
Matches 47; Conservative 46; Mismatches 51; Indels 8; Gaps 6;

26 PAEGTKSPKSRHRDRUNTELDRLASLLP--FPQ-DVIN-KLQKLSVLRSLVTYLRKSF 81
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
25 PQGUTCSGERRRQESKYTEELAEISANLSDIDFNVPKRCALLKRTVQRIQITE 94
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
82 FDVALKTPADRSQDQCRAQ-VRDQMDLQGEFLQALNGFVLVVTADALVFYASSRI 140
   ::|::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

```

```
DR EMBL; M55654; G339492; -.
DR EMBL; X54993; G37066; -.
DR PIR; A34830; A34830.
DR PIR; A34831; A34831.
DR PDB; 1TGH; 01-AUG-96.
DR PDB; 1CDW; 23-DEC-96.
DR MIM; 600075; -.
DR PROSITE; PS00351; TRIID; 2.
DR PFAM; PF00352; TBP; 2.
DR TRANSFAC; T00794; -.
KW TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN; REPEAT;
KW 3D-STRUCTURE.
FT DOMAIN 55 95 POLY-GLN.
FT REPEAT 165 241 1.
FT REPEAT 255 332 2.
FT CONFLICT 58 61 MISSING (IN REF. 2).
SQ SEQUENCE 339 AA; 37698 MW; 5D24A7B8 CRC32;

Query Match 2.3%; Score 225; DB 1; Length 339;
Best Local Similarity 38.5%; Pred. No. 1.17e-15;
Matches 42; Conservative 16; Mismatches 49; Indels 2; Gaps 2;

Db 18 QGAMTPGPIPSMPYGTG-LTPQPIQNTNSLLEQQRQQRQQRQQRQQRQQRQQRQQRQ 76
Qy 1196 ENPTAGGAVMRPMQPOQGFNAQVQRSR-ELLSHHFQQRVAMMMQQRQQRQQRQQR 1254
Db 77 QQQQQQQQQQQQQQQQQQAAVAAVQOSTQQATQGTSGQAPQLFHSQ 125
Qy 1255 QQQQQQQQQQQQQQQQQQATQAFSPNPNTASPMQGLLAGTPTMPQAPPQFPYQ 1303

RESULT 5
ID ARNT DROME STANDARD; PRT; 642 AA.
AC O15945; O16167;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR HOMOLOG (DARNT)
DE (TANGO PROTEIN).
GN TGO OR ARNT.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RC MEDLINE; 98040551.
OHSHIRO T., SAIGO K.;
"Transcriptional regulation of breathless FGF receptor gene by
binding of TRACHEALLESS/DARNT heterodimers to three central midline
elements in Drosophila developing trachea.";
RL DEVELOPMENT 124:3975-3986(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RC MEDLINE; 97427859.
RA ZELZER E., WAPPNER P., SHILO B.-Z.;
"the PAS domain confers target gene specificity of Drosophila
bHLH/PAS proteins.";
RT GENES DEV. 11:2079-2089(1997).
CC -1- FUNCTION: TGO/TRH HETERODIMERS ARE INVOLVED IN THE CONTROL OF
BREATHLESS EXPRESSION.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH AN OTHER
BHLH PROTEIN. HETERODIMER WITH AHR, TRH OR SIM.
CC -1- DEVELOPMENTAL STAGE: DISTRIBUTED HOMOGENEOUSLY IN EARLY EMBRYOS,
SUGGESTING MATERNAL DEPOSITION OF RNA. AT STAGE 11, EXPRESSION IS
DETECTED IN TRACHEAL PITS. AT LATER STAGES, STRONG EXPRESSION IS
ALSO DETECTED IN THE CNS.
CC -1- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
TRANSCRIPTION FACTORS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AB002556; D1023738; -.
DR EMBL; AF016053; G2367597; -.
DR FLYBASE; FBgn0015014; tgo.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
DR PFAM; PF00010; HLH; 1.
DR PFAM; PF00989; PAS; 1.
KW NUCLEAR PROTEIN; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW REPEAT.
FT DNA_BIND 14 26 BASIC DOMAIN.
FT DOMAIN 27 67 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT REPEAT 87 154 PAS-1.
FT REPEAT 273 339 PAS-2.
FT DOMAIN 346 389 PAC MOTIF.
FT DOMAIN 23 26 POLY-ARG.
FT DOMAIN 396 399 POLY-ALA.
FT DOMAIN 404 413 POLY-GLN.
FT DOMAIN 455 462 POLY-GLN.
FT DOMAIN 515 522 POLY-GLN.
FT DOMAIN 566 570 POLY-PRO.
FT CONFLICT 110 110 M -> V (IN REF. 2).
FT CONFLICT 411 411 Q -> QQQ (IN REF. 2).
FT CONFLICT 465 465 R -> G (IN REF. 2).
FT CONFLICT 488 488 P -> T (IN REF. 2).
FT CONFLICT 614 632 EFSDMLQMLDHTTFEDL -> GVLRYADVGSHADHV
(IN REF. 2).
SQ SEQUENCE 642 AA; 71606 MW; 983D5207 CRC32;

Query Match 2.3%; Score 224; DB 1; Length 642;
Best Local Similarity 34.5%; Pred. No. 1.67e-15;
Matches 50; Conservative 39; Mismatches 42; Indels 14; Gaps 8;

Db 19 CEIERRRNRKMTAYITELSDMVPTC-SALAR---KPKLTILRMVAHMKALRGNT-S 73
Qy 31 CSGEKRRREQESKYIELAELISANLSDIDNFNVKPKCALIKETVRIQIKQEGKTIS 90
Db 74 SDGIY-KP-SF---LTDQF-LAKHLILEADGFLFVVCDSGRWYVSDSVTPVLTQ 125
Qy 91 NDDVDQADVSSSTCGQVIDKSDLSGLGLLQALDGLFVNVNREAN-IVFVSENVTLQTKQ 149
Db 126 SDWYGTSLYEHHPDREKIREQLS 150
Qy 150 EDLVNTSVYNILHEEDRKDFLKNLP 174

RESULT 6
ID BTD DROME STANDARD; PRT; 644 AA.
AC Q24266;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION FACTOR BTD (BUTTONHEAD).
GN BTD.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S; TISSUE=EMBRYO;
RX MEDLINE; 94081952.
RA WIMMER E.A., JAECKLE H., PFEIFLE C., COHEN S.M.;
"A Drosophila homologue of human Spl is a head-specific segmentation
gene.";
RT NATURE 366:690-694(1993).
RL NATURE 366:690-694(1993).
CC -1- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF THE ANTENNA,
```


QY	1105	DQAK-LYGOTPYAQGPPMQ-GGFHLGQSPSFNSMMNOMN--OOGNFPLQ-GMHPRANI	1159
Db	454	FQQPM-MP-O-RTGLOPQMTGFGQPPVQRTGMPQMTGFGQPPMP-ORTGLOPQMTGF-	508
QY	1160	MRPTNTPKLRWLOLQRLQ-GOFLNQSQAELKWNPTAGGAAYVRPMQPPQ-GFL	1217
Db	509	QQPVPORTGMQPMGLQPPMAPORTGMQPPMPQRTGMQPPMTGFGQPPMAPQRTGMQPM	568
QY	1218	NAQVAQRS-RE-LLS--HH-FRQORVAM-MOQOQOQOQOQOQOQOQOQOQOQOQAF	1271
Db	569	MPORTGMQ-PQPMQMO-QPMAQRTGMQ-PMPQRTGM-QOPMAP-RTGMQPM--MMPQR	621
QY	1272	SPPNVTASPMGLLAGTMAPQAPQOFPYQPN-YGMQOQDPFAFGRVSSPPNAMSSR	1330
Db	622	TGMQPPMGQOQPMAPORTGMQPPMPQRTGMQPPMG-QOPMAP-QRTGMQPMAPQRTG	679
QY	1331	MG-PSQPMQHPQOASISYOSSEMKGWPSGNLARNSFSQOQFAGQNPAYISMV-HMNG	1388
Db	680	MQPMQPORTGMQ-QMPGMQOQMAPOR 705	
QY	1389	SSGHMGQMN-MNPMPSGM--PMGPDQ 1412	
RESULT 10			
ID	GDAL.WHEAT	STANDARD;	PRT; 262 AA.
AC	P04721;		
DT	13-AUG-1987 (REL. 05, CREATED)		
DT	13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)		
DT	01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)		
DE	ALPHA/BETA-GLIADIN PRECURSOR (PROLAMIN) (CLASS A-I).		
OS	TRITICUM AESTIVUM (WHEAT).		
CC	EUKARYOTA; VIRIDIPALTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;		
CC	EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;		
CC	POACEAE; TRITICUM.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 85234522.		
RA	OKITA T.W., CHEESBROUGH V., REEVES C.D.;		
RT	"Evolution and heterogeneity of the alpha-/beta-type and gamma-type		
RT	gliadin DNA sequences.";		
RL	J. BIOL. CHEM. 260:8203-8213(1985).		
CC	-1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.		
CC	-1- THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5 HOMOLOGY CLASSES.		
CC	SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO SINGLE BASE		
CC	SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS WITHIN OR NEAR		
CC	DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF THE GENE		
CC	FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.		

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		

DR	EMBL; M11074; G170722; -		
DR	PIR; B23364; B23364.		
KW	SEED STORAGE PROTEIN; REPEAT; SIGNAL; MULTIGENE FAMILY.		
FT	SIGNAL		
FT	1 20		
FT	CHAIN 21 262		
FT	SEQUENCE 262 AA; 30403 MW; 229A083D CRC32;		
QY	Query Match 2.2%; Score 219; DB 1; Length 262;		
QY	Best Local Similarity 31.6%; Pred. No. 9.70e-15;		
QY	Matches 50; Conservative 33; Mismatches 69; Indels 6; Gaps 6;		
Db	22 RVVPQLQPNPSSQOQEQVPLVQOQOFLQ-QOPFPQOQYPPQPPFPFSQOQVLOLQP 80		
QY	1156 RANIMRPTWPKLRWLOLQRL-QGOFLNQSQAELKWNPTAGGAAYVRPMQPPQ 1214		
Db	81 -FLQPLPYSPQPPFRPPQYPPQ-PQYSQPPQISQOQOQOQOQOQOQOQOQOQOQ 138		

QY 1300 FYPQNYGMOQDPAPGAFGRVSSPPNAMSRRMGPSQNMQMHPQAAAIYQSSSEMKGWPSG 1359
Db 584 AMNPLNHFHGHNNHHH 599
QY 1360 NLARNSFSQOQFAHQ 1375

RESULT 13
ID V192 HUMAN STANDARD; PRT; 2124 AA.
AC Q93074;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN KIAA0192 (FRAGMENT).
GN KIAA0192
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]

SEQUENCE FROM N.A.
TISSUE=BONE MARROW;
RX MEDLINE; 96281124.
RA NAGASE T., SEKI N., ISHIKAWA K.-I., TANAKA A., NOMURA N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RL analysis of cDNA clones from human cell line KG-1.";
RL DNA RES. 3:17-24(1996).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D83783; D1012786; -.
KW HYPOTHETICAL PROTEIN.
FT NON_TER 1
FT DOMAIN 599 602 POLY-SER.
FT DOMAIN 1201 1207 POLY-GLY.
FT DOMAIN 1998 2124 GLN-RICH.
FT DOMAIN 1998 2023 POLY-GLN.
FT DOMAIN 2028 2033 POLY-GLN.
FT DOMAIN 2037 2070 POLY-GLN.
FT DOMAIN 2090 2097 POLY-GLN.
SEQUENCE 2124 AA; 237207 MW; 24317D02 CRC32;

Query Match 2.28; Score 213; DB 1; Length 2124;
Best Local Similarity 34.78; Pred.No. 7.89e-14;
Matches 42; Conservative 23; Mismatches 55; Indels 1; Gaps 1;

Db 2003 QQQ 2062
QY 1211 QPQQGFLNAQVARSRELLSHFRQVRVAMMQQQQQQQQQQQQQQQQQQQQQQTQA 1270
Db 2063 HQQQQQQQAAPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 2122
QY 1271 FSPPPNVTASPSMDGLLAGTLPQAPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 1329
Db 2123 R 2123
QY 1330 R 1330

RESULT 14
ID HMCU_DROME STANDARD; PRT; 2175 AA.
AC P10150;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)

DE HOMEOBOX PROTEIN CUT.
GN CT.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE; 88232956.
RA BLOCHLINGER K., BODMER R., JACK J., JAN L.Y., JAN Y.N.;
RT "Primary structure and expression of a product from cut, a locus
RT involved in specifying sensory organ identity in Drosophila.";
RL NATURE 333:629-635(1988).
CC -!- FUNCTION: REGULATOR OF CELL FATE DECISIONS IN MULTIPLE LINEAGES.
CC SPECIFICALLY, FUNCTIONS AS A DETERMINATION FACTOR THAT SPECIFIES
CC SENSORY ORGAN IDENTITY IN PRECURSOR CELLS. PROBABLY ALSO INVOLVED
CC IN CELL TYPE SPECIFICATION OF MALPIGHIAN TUBULES. IN ABSENCE OF
CC CUT GENE EXTERNAL SENSORY ORGANS ARE TRANSFORMED INTO CHORDOTONAL
CC ORGANS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- TISSUE SPECIFICITY: DETECTED IN MANY CELLS IN THE CENTRAL NERVOUS
CC SYSTEM, ALL EXTERNAL SENSORY ORGANS, SOME PERIPHERAL NEURONS, AND
CC IN THE NON-NEURAL CELLS OF THE SPIRACLES AND THE MALPIGHIAN
CC TUBULES.
CC -!- DEVELOPMENTAL STAGE: CELL-SPECIFIC PATTERN OF EXPRESSION. BROADLY
CC EXPRESSED DURING EMBRYONIC DEVELOPMENT.
CC -!- DOMAIN: ASN AT POSITION 47 OF THE HOMEOBOX MAY PARTICIPATE IN
CC REGULATING DNA-BINDING ACTIVITY BY PROMOTING HOMO- AND
CC HETERODIMERIZATION.
CC -!- SIMILARITY: TO MAMMALIAN COAT DISPLACEMENT PROTEIN (CDP).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X07985; G7768; -.
KW HOMEOBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN;
KW COILED COIL; REPEAT.
FT DOMAIN 194 210 ALA/GLN-RICH.
FT DOMAIN 235 243 ALA-RICH.
FT DOMAIN 271 293 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 384 428 ASN-RICH.
FT DOMAIN 433 487 COILED COIL (POTENTIAL).
FT DOMAIN 547 554 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 574 584 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 616 630 ALA-RICH.
FT DOMAIN 665 699 HIS/GLN-RICH (OPA-REPEAT).
FT DOMAIN 886 1689 3 X 'CUT'-REPEATS.
FT REPEAT 886 958 'CUT'-REPEAT.
FT REPEAT 1339 1411 'CUT'-REPEAT.
FT REPEAT 1617 1689 'CUT'-REPEAT.
FT DNA_BIND 1745 1804 HOMEOBOX.
FT DOMAIN 2004 2014 ALA-RICH.
FT DOMAIN 2071 2077 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 2124 2136 ALA/PRO-RICH.
SQ SEQUENCE 2175 AA; 233628 MW; D7E1FC12 CRC32;

Query Match 2.28; Score 217; DB 1; Length 2175;
Best Local Similarity 23.18; Pred.No. 1.96e-14;
Matches 58; Conservative 77; Mismatches 99; Indels 17; Gaps 14;

This Page Blank (uspto)

QY 809 S--SDFYNNSIS-S-NGS---H--LGT-KQ---Q---VFQ-GTNSLGLKSSQSV-QSIRP 850
Db 876 GQLGRLLPNLPLDITLQSPGTGAGPEPPIRNSPVSVIPQGMWG-NOGMIGNOGLGN 934
QY 851 -----PY-N-R-AVLDSPVSGSSPPVRKISAFPLPKQPLGGMNPMDSQENTGS 900
Db 935 SGTGIGNSASRPTMPSGEWA-POSSAVRVTCATTSAMNRPV-Q-G-GMIRNP-AA5IP 989
QY 901 SMGGPNRVTVTQTFSSGDWGLPNSKACRME-PMNSNSMGRPGDYNTSLRPLAGG5IP 959
Db 990 ---MRPSSQPGQRTLQSV-N-NIGPSELENNMGPGPOYSQQQAPPNTATAPWPSILPID 1044
QY 960 TLPLRSNIPGARPLYOQQOQMLQRPGEIPMGMGANPYGA-AA5NQLG5WPDGML5ME 1018
Db 1045 QASFASQNRQPPGSPDCLCPHPAAE5P5DEGALLDOLYLALRFD--GLEEIDRALGI 1102
QY 1019 QVSHGTQNRPLRLN5DL5V5P5N5LEQ5DERALLDQLHLL5NT5DATG5LEEIDRALGI 1078
Db 1103 PELVSQ5QAVDPEQ--F5S0DSNIMLEQAPVFPQOYASQAQMAQGSYS-PMQDPNFHTM 1159
QY 1079 PELVNOGQALEPKDAFOGQEA5VMDQKAGLYGQTYPAQGP5PMQGGF5HLOGQ5P5FNSM 1138
Db 1160 -GQ--RPSYATLR-MQPRPG-LRPTGLVQNPQNLRLQLQHLRQAQ--N--RQPL-MNQ 1209
QY 1139 MNQMNQOGNFP5LQGMHPRANIMRPR--T-NTPKQLRMQLQORLQOQ5P5LQ5RQALELKM 1195
Db 1210 ISNV5NVL5LRPGV-PTQAPINAQMLAQROR5EILNHLRQ-R-QM-----HQ 1254
QY 1196 ENPTAGGA5V5M5P5Q5Q5G5F5L5Q5V5A5Q5R5L5H5F5Q5R5V5M5Q5Q5Q5Q5Q5Q5 1255
Db 1255 QQQVQOQRTLMRGQGLNMT5P5V5P5G5M5P5AT5N5P5RI5Q5A5Q5P5P5P5N5Y5G5IQ5Q5P5G 1314
QY 1256 QQQQQQQQQQQQQTQAF5P5PNV5T5P5SD5GL5LAG5T5M5P5Q5P5Q5P5Y5P5N5G5MG5Q5P5P5A 1315
Db 1315 FTGATTP5P5L5P5MAHT5P5QM5Q5Q5AN5P5AY5Q5P5D5ING5AQN5G5N5F5Q5Q5P5PH 1374
QY 1316 FQV55P5P5N5M55R5G5P5N5P5M5Q5H5Q5A55I5Q5-S5M5K5W5P5G5N5L5A5R5N5S5F5Q5---Q 1371
Db 1375 FQOQANT5M5N5M5N5Y5MAT5N5TG5M55N5QMT--GQISM 1413
QY 1372 FAHQ5NP5V5M--V5H5NG5-S5H5M5G5Q5M--M5P5M5P5M5G5M5P 1408

RESULT 4

ID Q61026 PRELIMINARY; PRY; 1462 AA.
AC Q61026; P97759;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GRIPI.
GN GRIPI.
OS MUS MUSCULUS (MOUSE).
OC EURARVOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
CC SCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RC STRAIN=ICR OUTBRED; TISSUE-BRAIN;
RP SEQUENCE OF 322-1119 FROM N.A.
RX MEDLINE; 96209838.
RA HONG H., KOHLI K., TRIVEDI A., JOHNSON D.L., STALLCUP M.R.;
RT "GRIPI, a novel mouse protein that serves as a transcriptional
coactivator in yeast for the hormone binding domains of steroid
receptors.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4948-4952(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR OUTBRED; TISSUE-BRAIN;
RX MEDLINE; 97265407.
RA HONG H., KOHLI K., GARABEDIAN M.J., STALLCUP M.R.;
RT domain of steroid, thyroid, retinoid, and vitamin D receptors.";
RL MOL. CELL. BIOL. 17:2735-2744(1997).
RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN=ICR OUTBRED; TISSUE-BRAIN;
RA STALLCUP M.R.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR OUTBRED; TISSUE-BRAIN;
RA HONG H., STALLCUP M.R.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EXBL; U39060; G1853980; --
DR MGD; MGI:107429; GRIPI.
DR PFAM; PF00989; PAS; 1.
SQ SEQUENCE 1462 AA; 158511 MW; 9F9EC6A7 CRC32;
Query Match 33.5%; Score 3300; DB 11; Length 1462;
Best Local Similarity 42.8%; Pred. No. 0.00e+00;
Matches 631; Conservative 375; Mismatches 339; Indels 129; Gaps 81;
Db 1 MSGGENTSDPSRAETRRK--ECPDQLGSPKPRSTERNRREQENKYEELADLIFANFD 59
QY 1 MSGLGENL-DPLASDRKRKLPC-DTPCQGLTCSGKRRRQESKYIEELAEELISANLSD 58
Db 60 IDNFNPKDCAILKETVKQIRQKEKQAAAAANIDEVQKSDVSTGGGVTDKDALGPM 119
QY 59 IDNFNPKDCAILKETVKQIRQKEQKQKTSND-DVQKADVSTGGGVTDKDSGLPLL 117
Db 120 LEALDGGFFVYNLEGSVYFVSENVTOYLRYNQEELMKSVY5ILHVGDHTEFVNKLPS 179
QY 118 LQALDGGFLVFNREANIVFVSENVTOYLQYKQEDLVNTSVNIIHEEDRKDFLKNL-PKS 176
Db 180 MVNGSGSGEPPRTSTFTNCRLMLVKPLPDSEEGHDSQEAHQKYEAMQCFAVSQPSIK 239
QY 177 TVNGSVTNEPQKSHFTNCRLMLKTPHDILEINASPENRQRYETMQCFAL5QPRAMM 236
Db 240 EGEDLQSLCIVARRVPMKERPTLPSS-ESFTTRDQLOGKITSLD5T5MRAAMKPGWED 298
QY 237 EGEDLQSLCIVARRITGER-TFSPNPE5FITRHLDSGKV5NIDT5NLSR5MRPGFD 295
Db 299 LVRCIOKFTQHEGESLSYAKRHHHEVLROGLAF5IYREF5LSDGTLVAQTKSLIRS 358
QY 296 IIRCIORFFSLNDGQ5WS-QKRHYEAYLNGHAETPVYREF5LADG5TIVTAQTKSLFRN 354
Db 359 QTTNEPQIV5LHMLHREQNV5VMNPDLT5QAMGKPLNP5SSSP5HAQ5L5G5NPGQMT 418
QY 355 PVTNDRHGFVSTHFLQREQNYRPNPNVQGI-RP--PWAGCN5SVGGM-SM5PNOGLQ 410
Db 419 LGSNINFP5MNGPK5QM5M5P5G5G5G5M5N5V5--Q5QATTP-Q5G5N5YALKM5SP5Q5S 474
QY 411 MP5SRAYGLADP5T5TG5M5G5ARY5G55N5IASLT5PG5M5Q5P55Y5Q5N5N5YGLM5SP5PHG5 470
Db 475 PGMNPGQASSV5L5PRQ5M5PGV5G5PRIP5Q5F5P5AG5L5H5P5G5V5C5T5G5N5H5Y5T5N5SL 534
QY 471 PGLAPNQNIM5IPNR-----G5P5IASHQ5F5P5V5AG5H5P5MA55NGT5GN-H5F5SS5L 523
Db 535 NALQAL5EGH5V5L5G55L5AS5DL5K5M5GL5Q5N5P5M5P5P5L5K5M5GL5DK5C5F5GLY-GE-P 592
QY 524 SALQAISEG5V5T5L5L5T5L5SP5G5PK--LDNSP-NMNITQ5P5K5V5N5Q5D5K5PL5G5Y5C5QNP 579
Db 593 SEGTTQ5A5E5A5CH5PE5Q5K5P5ND5SM--PQA5GD-R5E5GH5RL-H--DSK5Q5TK5LL5QLLT 646
QY 580 VESSM5Q5N5SRD5HL5D-KESKES5VEGA5N5Q5R5P5L5K5GH5K5KL5QLLT5C5SD5DR5H5SLT 638
Db 647 TKS-DQM-EP5PLP-SSISD5TNK5D5T5G5L5P5G5T5H5C5T5L5K5E5KH5K5HL5R5L5Q5D5SS5VDL 703
QY 539 NSPLD5CK5ESS5V5T5P5G5V5SS5T5G5V5ST5N5H5G5SL5Q5E5KH5R5L5K5L5Q5N55PA5E5 698
Db 704 AKLTAEATG5L5Q5E5SS5T5AP5G5E5TV5K5Q5E5P5ASP5PK5KEN-ALL5RYL5DK5D5TK5D5IGL5P-E 761
QY 699 AKITAQATG5D5T5SIT5SCGD-GN-V-VK5Q5L5SP5KK5ENALL5RYL5DR5D5PSD-ALSKE 754
Db 762 ITPK5L5R5D5SK5D5PAS5TK5LIAM5TV5E-EV5F5P5Q5P5Q5E5L5N5E5IL5D5L5Q5N5Q5LPQ 820
QY 755 LQ5QV5G5V5DN5KM5SQ5CT5STIP5SS5Q5E5K5D5PK5IK5T5E5G5D5L5N5D5AIL5D5L-TS--SD 811

D	b	1	MMDQKAGLGYQTYPAQGPPMQGGFHLCQSSFSFMSNMNQNGNPFPLQGMHPHINMRP	60
Q	y	1103	MMDOKAGLGYQTYPAQGPMPQGGFHLCQSSFSFMSNMNQNGNPFPLQGMHPHINMRP	1162
D	b	61	RTNTPQLRMQLQQRLLQGOQFLNQSRQALELKMNPTAGGAAVYRPMQMOPVQSQSGFLN	120
Q	y	1163	RTNTPQLRMQLQQRLLQGOQFLNQSRQALELKMNPTAGGAAVYRPMQMOPQ--Q-GFLN	1218
D	b	121	AQMGAGRRELLSHHFROQRVAMMOCQQCQCQCQCQCQCQCQCQCQCQCQCQCQCQ	180
Q	y	1219	AQMVAGRRELLSHHFROQRVAMMOCQQCQCQCQCQCQCQCQCQCQCQCQCQCQ	1272
D	b	181	PPNVNTAPANWAXGETHNAISSATSTQYP-TPNYNGDOPTKIPLGRVKSSOCNDV	239
			: : : : : : : : : :	
Q	y	1273	PPNVNTASPS-MD-GLLAGTTPMQAPPQFFPYQNYGMGQPD--PAFGRVSPPPNAMMS	1328
D	b	240	VKNPGSQNPMMQHQAASITQSEMKMGWPSCGNLARNSSFQQQFAHQGNPATVZMYVMHG	299
			:::	

Db	1	SGLGSSDDPANPDSHRRKSPCDT----LASTERRRREQENKYEEELAEELISANISDI	56
Qy	2	SGLGENT-LPLADSRARRK-LPDCDTPGGTCTGSEKRRREQESKYTEELAEELISANISDI	59
Db	57	DSLGVKPKCKILKTKTVDQQLMKRMEQKSTT-DDDVQKSDTSSSSQGVIEKSLGPLL	115
Qy	60	DNFNVKPKCALLKETVYRQIRK--EQGKTISSNDDDVQKADYSSVTCQGVIDRDSLGPLL	117
Db	116	LEALDGEFFVYVNCBGRIVFVSENVTSYLVGNQBELMNTSVYSILHVGDDHAEFVKNLPPKS	175
Qy	118	LQALDGLFLVYNREANIVFVSENVTYQLYKQBDLVNTSVYNTLHEEDRKDFLKNL-PKS	176
Db	176	LVNGVVPQEQATRNSHTFNCRLIHPP-D--EPGTENQACACQYEVMOCTFVSQPSKIQ	232
Qy	177	TVNGSVNTNPEQKQSHTFNCRMLMKTPHDLIEDIINASPEMRQRIETMOCFAULSQPRAM	236
Db	233	EDGEDFQSLCITAIARRLP-R-PAI-TGVESFMTKQDTTGKIIISDTSLSRAAGRTGWEDL	290
Qy	237	EEGEDLQSCMICVARRITTCGERTFPNSPEFIRPHDLGKGVINDTNSLRMSRWGPFEDI	296

291	VRKCIYAFQPGQREPSYARQLFEQVMTGRTASSPYRETLNDGTMLSAHTKCL-CYQP	349
297	IRRCIQREFSLNDQGSQSKRRHYQEAYLNCHAEPTVYREFSLADGTIVTAQTKSKLFRNPV	356
350	SPDQMPQIMGIHIDRPHSLFQDDDTNSCKMTPRVNPSVNPSPISPAHGVARSTLPPSN	409
357	TNDRHGKV-STHFLQREQNGYRPNPNP-VQGGI-R--PPMAGCNSV-G-GMSMS-PNOG	408
410	SNMYSTRINQQSSDDLHSSSHSNSSNSQSGFGSPGSGQIVANVALNQOGASSQSNPSLN	469
409	LQMFSSRAY-GLA-DPSTTQCMGARYGGS-S-NIAS-LTPGFCM-QSPSSYONNNYGLN	462
470	LNNSPMGTGTLISLAQFMSPRRQVTSGLATRPMPNNSFPNPITLSTSPVGMTSSACNNNN	529
463	MSSSP-HGSP-GLAP-NQQNIMISPRNRGSPKTAHQFSP-VAGVHSP--MASSG-NTGN	515
530	RSYSNIPVTSIQMGNECPNNSVGFSASSPVLROMSSQNSPSRLNI-QPAKAESKDINKETA	588
516	HSFSSSLALQAISEGVGTSLLSTLSSPGPK-L--DNSPN-WNITOPSKVSNQDSKSP	571
589	SIL-NEM-IQSDNSSDGK-PL-DSGLLHNDRSLDGD-SKYSQTSHK-LVQLLTT-TTAE	641


```

RESULT 11
ID Q61202; PRELIMINARY; PRT: 1405 AA.
AC Q61202;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE NUCLEAR RECEPTOR COACTIVATOR 1 (STEROID RECEPTOR COACTIVATOR 1A).
GN NCOAL OR SRCIA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA KAMEI Y., XU L., HEINZEL T., TORCHIA J., KUROKAWA R., GLOSS B.,
RL LIN S.C., HEYMAN R.A., ROSE D.W., GLASS C.K., ROSENFELD M.G.;
RL CELL 0:0-0:0.
DR EMBL; U56920; G1336160; -.
DR MGD; MGI:1276523; NCOAL.
DR PFAM; PF00989; PAS; 1.
SEQUENCE 1405 AA; 152571 MW; 670CD071 CRC32;

Query Match 16.9%; Score 1665; DB 11; Length 1405;
Best Local Similarity 39.7%; Pred. No. 2,05e-280;
Matches 338; Conservative 229; Mismatches 215; Indels 70; Gaps 55;

Db 1 MSGIGSSSDPANPDHSHKRGKSPCDT----LASTKRRRQENKYLEGLAELLSANISD 56
QY 1 MSGIGENL-DPLASDSKRK-LPCDTPGQGLTCGKRRRQESKYIEELAEELISANLSD 58
Db 57 IDSLSVKPKCKTLKKTVDQIQLMKRMEQKSTT-DDDVOKSDISSSQGVIEKESLGPL 115
QY 59 IDNFNVPKPKCALIKETVROIRK--EOGKTTISNDVQKADVSTGQGVIDKDSLGLPL 116
Db 116 LLEALDGFVFNVECEGRIVFVSENVTSYLYGNOEELMNTSVYILHVGDAEFAVKNLKP 175
QY 117 LLEALDGFVFNVECEGRIVFVSENVTSYLYGNOEELMNTSVYILHVGDAEFAVKNLKP 175
Db 176 SLVNGVWPQEAATRRNSHTFNCRLIHPPED--PGTE-NOEAQRYEVMOQFTVSPQSKI 232
QY 176 STVNGVSWTNEPQOKSHTEFNCRLMKTPHDILEDINASPEMRQRYETMQCFALSOPRAM 235
Db 233 QEDGDFQSLCIAARLPP-PAI-TGVESFMTKQDTTKIISIDTSSRAAGRTGWED 290
QY 236 MERGEDLQSCMICVARRITTGERTFSPNPFITRDLGKGVNIDTNSLRSSMRPGFED 295
Db 291 LVKRCIYAFPOQREPSYAROLFQEVMTGTASSPSYRFLINDGTMLSAHTCKKL-CYP 349
QY 296 IIRRCIORFSLNDGQSWQKRYQEAYLNGHAETPVYRFLADGTLIVTAQTKSLFRNP 355
Db 350 QSPDMQPFIMGIHIDREHSGLSQDDNSGMSIPRINPSVNPFGISPAHGVTRSTLPPS 409
QY 356 VTNDRHGFV-STHFLQREQNGYRPNP-VQOGI-R--PPM-AGCNSSVG-GMSMS-PNQ 407
Db 410 NNMVSVARVNRQSSDLNSSSHTSNNOGNGFCGPGNOIVANVALNQOAGSQSTNPS 469
QY 408 GLOMPSSRAY-GLA-D-PSTTGOMSGARY-G--GSSNIASTLTPGGM-QSPSSYQNNYNG 460
Db 470 LNLNNSPMEGTGALSOFMSPRQANSGLATRAMSNNSFPNPTLSSPVGITSACNN 529
QY 461 LNMSSPPHSGPLAPNOQNTIMISPRNG--S-PKIAHQFSP-VAGVHSPMA--SSG--NT 513
Db 530 NNRYSNIPVTSLQGMNPGNNV-G-FSAGSPVLRQMSQN-S-PSRLSQPAKAESKD 585
QY 514 GNHSFSSSLSAQAISEGVGTSLTSLSPGPKLDNSPNNITQPKSVNQDSKSLGF 573
Db 586 SKETASILNMIQSDNDN-SANGKPLDGLLHNDRLSEGDSKYQTSKHLVQLLT-T 643
QY 574 YCDQNPVSSMCQNSRDHLSDKESKE--SSVEGAENQR-GPLESK-GHK-KLIQLLTC 628
Db 644 TAEQ-Q--LRHADITCKDV-LSCTGTSSASNPSPGCTCPSS-HSSL-TERKHILHRL 697
QY 629 SDDRGHSSLTNSPLDSSCKESSVTSPPSGVSSSTSGVSTSNMHGSLLOEKHRIHLK 688

```

```

Db 698 IQEG-SPSDITLTSVEPEKDSVPASTAYSVSQSGSASIKLELDAAKKESKHQLLR 756
QY 689 LONGNSPAEVAKITAQTGKDT--SSIT-S-CGD--GNV-VKQEQSLPKKKE--NNALLR 739
Db 757 YLLDKDEKDLRSTPNCLDDVKVKEKQMDPCNTNPTMTKPAPE-EVKLESQSOFTA 815
QY 740 YLLDRDDPSALSKEL-QPOVEG-VDNK--MSQCTSTTSPSSQEKDPKIKETSEGG 795
Db 816 DLDQFQDQLLPTL 827
QY 796 DLDNLDAILGDL 807

RESULT 12
ID P70365; PRELIMINARY; PRT: 1447 AA.
AC P70365;
DT 01-FEB-1997 (TREMREL. 02, CREATED)
DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE NUCLEAR RECEPTOR CO-ACTIVATOR.
GN MNRG-1.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ZHU Y., QI C., CALANDRA C., RAO M.S., REDDY J.K.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U64606; G1490872; -.
DR PFAM; PF00989; PAS; 1.
SEQUENCE 1447 AA; 156896 MW; A92E6387 CRC32;

Query Match 16.7%; Score 1642; DB 11; Length 1447;
Best Local Similarity 39.4%; Pred. No. 5,31e-276;
Matches 336; Conservative 231; Mismatches 215; Indels 70; Gaps 55;

Db 1 MSGIGSSSDPANPDHSHKRGKSPCDT----LASTKRRRQENKYLEELAEELLSANISD 56
QY 1 MSGIGENL-DPLASDSKRK-LPCDTPGQGLTCGKRRRQESKYIEELAEELISANLSD 58
Db 57 IDSLSVKPKCKTLKKTVDQIQLMKRMEQKSTT-DDDVOKSDISSSQGVIEKESLGPL 115
QY 59 IDNFNVPKPKCALIKETVROIRK--EOGKTTISNDVQKADVSTGQGVIDKDSLGLPL 116
Db 116 LLEALDGFVFNVECEGRIVFVSENVTSYLYGNOEELMNTSVYILHVGDAEFAVKNLKP 175
QY 117 LLEALDGFVFNVECEGRIVFVSENVTSYLYGNOEELMNTSVYILHVGDAEFAVKNLKP 175
Db 176 SLVNGVWPQEAATRRNSHTFNCRLIHPPED--PGTE-NOEAQRYEVMOQFTVSPQSKI 232
QY 176 STVNGVSWTNEPQOKSHTEFNCRLMKTPHDILEDINASPEMRQRYETMQCFALSOPRAM 235
Db 233 QEDGDFQSLCIAARLPP-PAI-TGVESFMTKQDTTKIISIDTSSRAAGRTGWED 290
QY 236 MERGEDLQSCMICVARRITTGERTFSPNPFITRDLGKGVNIDTNSLRSSMRPGFED 295
Db 291 LVKRCIYAFPOQREPSYAROLFQEVMTGTASSPSYRFLINDGTMLSAHTCKKL-CYP 349
QY 296 IIRRCIORFSLNDGQSWQKRYQEAYLNGHAETPVYRFLADGTLIVTAQTKSLFRNP 355
Db 350 QSPDMQPFIMGIHIDREHSGLSQDDNSGMSIPRINPSVNPFGISPAHGVTRSTLPPS 409
QY 356 VTNDRHGFV-STHFLQREQNGYRPNP-VQOGI-R--PPM-AGCNSSVG-GMSMS-PNQ 407
Db 410 NNMVSVARVNRQSSDLNSSSHTSNNOGNGFCGPGNOIVANVALNQOAGSQSNPS 469
QY 408 GLOMPSSRAY-GLA-D-PSTTGOMSGARY-G--GSSNIASTLTPGGM-QSPSSYQNNYNG 460
Db 470 LNLNNSPMEGTGALSOFMSPRQANSGLATRAMSNNSFPNPTLSSPVGITSACNN 529
QY 461 LNMSSPPHSGPLAPNOQNTIMISPRNG--S-PKIAHQFSP-VAGVHSPMA--SSG--NT 513
Db 530 NNRYSNIPVTSLQGMNPGNNV-G-FSAGSPVLRQMSQN-S-PSRLSQPAKAESKD 585

```


OY 1288 AGPTMPQAPPQOPYPQPNYGMGQDPDPAFGRVSSPPNMMSSRMGP-SON-PMQHPOA 1344

Search completed: Fri Sep 17 21:16:25 1999
Job time : 245 secs.

This Page Blank (uspto)

QY	3768	AGTGTCTAGTCAATCATCTTCCGACACACAGAGAGGTGGCTATGATGATGCAGCAGCAGCAAC	3827
Db	44	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	103
QY	3828	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	3887
Db	104	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	163
QY	3888	CCCAGGCCTTCAGCCCAACCTCC	3909


```

: TELEPHONE: 612-305-1217
:
: TELEFAX: 612-305-1228
:
: INFORMATION FOR SEQ ID NO: 6
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 154 base pairs
:
: TYPE: nucleic acid
:
: STRANDEDNESS: single
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA
:
: US-08-267-803B-6

```

```
Query Match      1.5%; Score 68.4; DB 4; Length 154;
Best Local Similarity 71.4%; Pred. No. 3.3e-10;
Matches 90; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```

			GAGCTGCTAAGTCATCATCCGCACACAGAGGGTGGGTATTGATGATCGCAGCAGCAA	3876
OY				
			2 GAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	61
Dd				
			CAGCAGCAGCAGCAGCAGCAGCAACAGCACAGCAGCAGCAGCAGCAGCAGCAA	3886
OY				
P			62 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	121

QY 3887 ACCCAG 3892
|||
Db 122 CAGCAG 127

```

RESULT 14
US-08-267-803B-7
: Sequence 7, Application US/08267803B
: Patent No. 5834183
: GENERAL INFORMATION:
: APPLICANT: Ori, Harry T.
: APPLICANT: Rarum, Laura P.W.
: APPLICANT: Chung, Ming-yi
: APPLICANT: Zoghbi, Huda Y.
: TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
: Patent No. 5834183
: TITLE OF INVENTION: Type 1 and Method for Diagnosis
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Muetling, Raasch, Gebhardt & Schwappach, P.A.
: STREET: P.O. Box 581415
: CITY: Minneapolis
: STATE: MN
: COUNTRY: USA
: ZIP: 55458-1415

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/267,803B
FILING DATE: 28-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McCoimack, Myra H.
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110,00030120
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 506 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-267-803B-7

```

US-08-267-803B-7

	Query Match	1.5%	Score 68.4;	DB 4;	Length 506;
	Best Local Similarity	66.0%	Pred. No. 6.7e-10;		
	Matches 99;	Conservative 0;	Mismatches 51;	Indels 0;	Gaps 0;
QY	3760	CAGCAGAGAGCTGTAAGTCATCACTTCCGACACAGAGGGTGGCTATGATGATGCAGCA	3819		
Db	105	CAACATGGGAGTCTGAGCCAGAGCGCGGACACAGAGGCTGAGCAGCAGCAGCAGCA	164		
QY	3820	GCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAACAGCAACAGCAGCAACA	3879		
Db	165	GCAGCAGCAGCAGCAGCAGCATCAGCATCAGCAGCAGCAGCAGCAGCAGCAGCA	224		
QY	3880	GCAGCAAAACCCAGCCCTTCAGCCCAACCTCC	3909		
Db	225	GCAGCAGCAGCAGCAGCAGCCTCAGCAGGGGTCC	254		

```

RESULT 15
US-08-267-803B-8
; Sequence 8, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Ori, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueeting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415

```

```

, , COMPUTER READABLE FORM:
, , MEDIUM TYPE: Floppy disk
, , COMPUTER: IBM PC compatible
, , OPERATING SYSTEM: PC-DOS/MS-DOS
, , SOFTWARE: PatentIn Release #1.0, Version #1.25
, , CURRENT APPLICATION DATA:
, , APPLICATION NUMBER: US/08/267,803B
, , FILING DATE: 28-JUN-1994
, , CLASSIFICATION: 435
, , ATTORNEY/AGENT INFORMATION:
, , NAME: McCormack, Myra H.
, , REGISTRATION NUMBER: 36,502
, , REFERENCE/DOCKET NUMBER: 110.00030120
, , TELECOMMUNICATION INFORMATION:
, , TELEPHONE: 612-305-1217
, , TELEFAX: 612-305-1228
, , INFORMATION FOR SEQ ID NO: 8:
, , SEQUENCE CHARACTERISTICS:
, , LENGTH: 10660 base pairs
, , TYPE: nucleic acid
, , STRANDEDNESS: single
, , TOPOLOGY: linear
, , MOLECULE TYPE: DNA
, , FEATURE:
, , NAME/KEY: CDS
, , LOCATION: 936..3384
, , US-08-267-803B-8

```

Query Match 1.5%; Score 68.4; DB 4; Length 10660;
Best Local Similarity 66.0%; Pred. No. 4.2e-09;
Matches 99; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 3760 CAGCAGAGGCTGTTAGTTCATCTACTCCGACACAGAGGGTGGCTATGATGATGCACGA 3819

Result No.	Query	Score	Query			ID	Description
			Match	Length	DB		
1	4451.6	99.0	6835	1	V99915	AIB1 Amplified in	
2	2812.2	62.5	4621	1	V99919	Murine pCIP gene (
3	2750.8	61.2	4860	1	X26000	Murine pCIP gene.	
4	561	12.5	6156	1	V03517	Human transcriptio	
5	432.4	9.6	462	1	V44857	Clone AJ1_1 5' cod	
6	392.4	8.7	418	1	V90274	EST clone D1454. N	
7	261.8	5.8	325	1	V86262	EST clone AJ1. New	
8	76.2	1.7	1310	1	V08558	Transgene for epit	
9	76.2	1.7	4286	1	V08559	Transgene for epit	
10	76.2	1.7	3263	1	V08560	Transgene for epit	
11	73.2	1.6	195	1	V084831	Spino cerebellar at	
12	73.2	1.6	234	1	V084832	Spino cerebellar at	
13	73.2	1.6	168	1	V084833	Spino cerebellar at	
14	73.2	1.6	171	1	V084834	Spino cerebellar at	
15	68.6	1.5	645	1	T42476	HuntL4 coding sequ	
16	68.6	1.5	2294	1	V80187	Alternatively splic	
17	68.6	1.5	3334	1	V80117	osf2/cbfa1 polypep	
18	68.4	1.5	3366	1	V84803	Spino cerebellar at	
19	68.4	1.5	506	1	V084804	Spino cerebellar at	
20	68.4	1.5	10660	1	V084793	Spino cerebellar at	
21	68.4	1.5	154	1	V084835	Spino cerebellar at	
22	67.8	1.5	203	1	V17226	SCA2 gene CAG repe	
23	67.8	1.5	203	1	V03271	Glutamine rich reg	
24	67.6	1.5	10366	1	V67401	Huntingtin DNA/It1	
25	67.6	1.5	10366	1	V97924	Composite sequence	
26	67.6	1.5	10366	1	V05828	Human huntingtin c	
27	67.6	1.5	515	1	X18930	Human huntingtin g	
28	67.4	1.5	515	1	V18930	Glutamine rich reg	
29	66.8	1.5	4367	1	V30270	Gene causative of	
30	66.6	1.5	1065	1	V74149	Plasmid pSI4001 al	
31	66.6	1.5	1680	1	Q74150	Plasmid pSI4001 al	
32	66.4	1.5	3376	1	V75166	AF-9 cDNA. New acu	
33	65.4	1.5	4200	1	V78912	Spino cerebellar at	
34	65.4	1.5	516	1	V08551	SCA2 gene fragment	
35	65.4	1.5	4481	1	V08552	Human SCA2 cDNA in	
36	64	1.4	379	1	Q04602	Cryptically simple	
37	62.8	1.4	3569	1	N91772	Human androgen rec	
38	62.8	1.4	3569	1	T63407	Androgen receptor	
39	62.6	1.4	2069	1	X22940	Mouse calpain CAPN	
40	62.4	1.4	3217	1	N91773	Rat androgen recep	
41	62.4	1.4	4180	1	N91778	Rat androgen recep	
42	62.4	1.4	3217	1	Q12002	Full-length rat an	
43	62.2	1.4	78	1	V17232	SCA2 gene CAG repe	

Db 356 ATCTGCCAATCTTAGTGATATTGACAAATTTCAATGTCAAACCCAGATAAAATGTGCGATTTT 415
Qy 301 AAAGGAACAGTAGACAGATAGCTCAAAATAAAGACCAAGGAAAACACTATTTCCCAATGA 360
Db 416 AAAGGAACAGTAGACAGATAGCTCAAAATAAAGACCAAGGAAAACACTATTTCCCAATGA 475
Qy 361 TGATGATGTTCAAAAACCCGATGATCTCTACAGGCGAGGAGTTATTGATAAAGACTC 420
Db 476 TGATGATGTTCAAAAACCCGATGATCTCTACAGGCGAGGAGTTATTGATAAAGACTC 535
Qy 421 CTTAGGACCGCTTTTACTTCAGGCATTGGATGGTTTCCCTATTGTGGTGAAATCGAGAGC 480
Db 536 CTTAGGACCGCTTTTACTTCAGGCATTGGATGGTTTCCCTATTGTGGTGAAATCGAGAGC 595
Qy 481 AATCATGTTATGTAACAGAAATGTCACACAATCTGCAACAATCTGCAATATAAGCAAGGACCT 540
Db 596 AATCATGTTATGTAACAGAAATGTCACACAATCTGCAACAATCTGCAATATAAGCAAGGACCT 655
Qy 541 GGTAAACACAAAGTGTTCACAAATCTTACATGAAGAAGACAGAAAGGATTTTCTTAAGAA 600
Db 656 GGTAAACACAAAGTGTTCACAAATCTTACATGAAGAAGACAGAAAGGATTTTCTTAAGAA 715
Qy 601 TTTACCAAAATCTACAGTTAATGGAGTTTCTGGACAAATGAGCCCCCAAGACAAAAG 660
Db 716 TTTACCAAAATCTACAGTTAATGGAGTTTCTGGACAAATGAGCCCCCAAGACAAAAG 775
Qy 661 CCATACATTAATTTGCCGTATGTTGATGAACACACACATGATATCTGGAAGACATAAA 720
Db 776 CCATACATTAATTTGCCGTATGTTGATGAACACACACATGATATCTGGAAGACATAAA 835
Qy 721 CGCCAGTCTGGAATGCCAGAGATATGAACAATGCAGTGTGTTGCCCTGTCTCAGCC 780
Db 836 CGCCAGTCTGGAATGCCAGAGATATGAACAATGCAGTGTGTTGCCCTGTCTCAGCC 895
Qy 781 ACGAGCTATGATGAGGAAGGGAAGATTGCAATCTTGTATGATCTGTGGCAGCCCG 840
Db 896 ACGAGCTATGATGAGGAAGGGAAGATTGCAATCTTGTATGATCTGTGGCAGCCCG 955
Qy 841 CATTACTACAGGAAGAAACATTTCCATCAAAACCTGAGAGCTTTATTACCAACAGATGA 900
Db 956 CATTACTACAGGAAGAAACATTTCCATCAAAACCTGAGAGCTTTATTACCAACAGATGA 1015
Qy 901 TCTTTACGAAAGGTTGTCAATATAGATACAAATTCACAGATCTCCCATGAGCCCTGG 960
Db 1016 TCTTTACGAAAGGTTGTCAATATAGATACAAATTCACAGATCTCCCATGAGCCCTGG 1075
Qy 961 CTTTGAAGATATAATCCGAAGGTGATTACAGAGATTTTGTAGTCTAAATGATGGCAGTC 1020
Db 1076 CTTTGAAGATATAATCCGAAGGTGATTACAGAGATTTTGTAGTCTAAATGATGGCAGTC 1135
Qy 1021 ATGTTCCAGAAAGCTCACATCAAGAAGCTTATCTTAATGGCCATCGAANAACCCAGT 1080
Db 1136 ATGTTCCAGAAAGCTCACATCAAGAAGCTTATCTTAATGGCCATCGAANAACCCAGT 1195
Qy 1081 ATATCGATTTCTGTTGGCTGATGGAATATAGTGACTGCAACAAAAGCAAACTCTTT 1140
Db 1196 ATATCGATTTCTGTTGGCTGATGGAATATAGTGACTGCAACAAAAGCAAACTCTTT 1255
Qy 1141 CCGAAATCTTGTAACAAATGATCGACATGGTTTGTCTCAACCCACTTCTTCCAGAGAGA 1200
Db 1256 CCGAAATCTTGTAACAAATGATCGACATGGTTTGTCTCAACCCACTTCTTCCAGAGAGA 1315
Qy 1201 ACAGAATGGATATAGACCAACCAATCTGTTGGACAGGATAGACCACTTATGGC 1260
Db 1316 ACAGAATGGATATAGACCAACCAATCTGTTGGACAGGATAGACCACTTATGGC 1375
Qy 1261 TGGATGCAACAGTTTCGGTAGCGGCATGAGTATGTCGCCAAAACCAAGCTTTACAGATGCC 1320
Db 1376 TGGATGCAACAGTTTCGGTAGCGGCATGAGTATGTCGCCAAAACCAAGCTTTACAGATGCC 1435
Qy 1321 GAGCAGAGGCGCTTATGGCTTGGCAGACCCCTAGCACACAGGCGAGATGATGAGGCTAG 1380
Db 1436 GAGCAGAGGCGCTTATGGCTTGGCAGACCCCTAGCACACAGGCGAGATGATGAGGCTAG 1495

Qy 1381 GTATGGGGTTCAGTAACATAGCTTTCATTGACCCCTGGCCAGGCATGCAATCACCATC 1440
Db 1496 GTATGGGGTTCAGTAACATAGCTTTCATTGACCCCTGGCCAGGCATGCAATCACCATC 1555
Qy 1441 TTCTTACCAGAACCAACTATGGGCTCAACATGAGTAGCCCCCACATGGGAGTCTCTGG 1500
Db 1556 TTCTTACCAGAACCAACTATGGGCTCAACATGAGTAGCCCCCACATGGGAGTCTCTGG 1615
Qy 1501 TCTTGGCCCAACACAGAGATAATCATGATTTCTCTCTGTAATCGTGGGAGTCCAAAGAT 1560
Db 1616 TCTTGGCCCAACACAGAGATAATCATGATTTCTCTCTGTAATCGTGGGAGTCCAAAGAT 1675
Qy 1561 AGCTCACATCAGTTTCTCTCTGTCAGGTGCACTCTCCCATGCACTCTTCTGSCAA 1620
Db 1676 AGCTCACATCAGTTTCTCTCTGTCAGGTGCACTCTCCCATGCACTCTTCTGSCAA 1735
Qy 1621 TACTGGAAACACAGCTTTTCCAGCAGCTCTCTCAGTGGCCCTGCAAGCCATCAGTGAAG 1680
Db 1736 TACTGGAAACACAGCTTTTCCAGCAGCTCTCTCAGTGGCCCTGCAAGCCATCAGTGAAG 1795
Qy 1681 TGTGGGACTTCCCTTTTATCTACTCTGTCATCACCAGGCCCCCAATTTGGATTAACCTCTCC 1740
Db 1796 TGTGGGACTTCCCTTTTATCTACTCTGTCATCACCAGGCCCCCAATTTGGATTAACCTCTCC 1855
Qy 1741 CAATATGAATATTACCAACCAAGTAAAGTAAGCAATCAGGATTTCCAAGAGTCTCTCTGG 1800
Db 1856 CAATATGAATATTACCAACCAAGTAAAGTAAGCAATCAGGATTTCCAAGAGTCTCTCTGG 1915
Qy 1801 CTTTATTGGACCAAAATCCAGTGGAGAGTTCAATTTGTCTCAGTCAAAATAGCAGAGATCA 1860
Db 1916 CTTTATTGGACCAAAATCCAGTGGAGAGTTCAATTTGTCTCAGTCAAAATAGCAGAGATCA 1975
Qy 1861 CTTCACTGACAAAGAAAGTAAGGAGAGCAGTGTGAGGGGGCAGAGAAATCAAAAGGGTCC 1920
Db 1976 CTTCACTGACAAAGAAAGTAAGGAGAGCAGTGTGAGGGGGCAGAGAAATCAAAAGGGTCC 2035
Qy 1921 TTTGGAAGCAAAAGGTCATAAAAAATTTACTGCAAGTACTTTACCTGTTCTCTCATGACCG 1980
Db 2036 TTTGGAAGCAAAAGGTCATAAAAAATTTACTGCAAGTACTTTACCTGTTCTCTCATGACCG 2095
Qy 1981 GGCTCATCTCTCTTGGACCAACTCCCCCTAGATTCAAGTTGTAAAGAAATCTTCTGTAG 2040
Db 2096 GGCTCATCTCTCTTGGACCAACTCCCCCTAGATTCAAGTTGTAAAGAAATCTTCTGTAG 2155
Qy 2041 TGTCAACAGCCCTCTGGAGTCTCTCTCTCTCATCTCGAGGAGTATCTCTACATCCAA 2100
Db 2156 TGTCAACAGCCCTCTGGAGTCTCTCTCTCTCATCTCGAGGAGTATCTCTACATCCAA 2215
Qy 2101 TATGCAATGGTCACTGTTCAAGAGAGCAGCCGATTTTGCACAAGTTGCTGCAGAGTGG 2160
Db 2216 TATGCAATGGTCACTGTTCAAGAGAGCAGCCGATTTTGCACAAGTTGCTGCAGAGTGG 2275
Qy 2161 GAATTCACCAAGCTGAGTAGCCAAAGATTACTGCACAAGCCACTGGGAAAGACACAGCAG 2220
Db 2276 GAATTCACCAAGCTGAGTAGCCAAAGATTACTGCACAAGCCACTGGGAAAGACACAGCAG 2335
Qy 2221 TATAACTTCTTGTGGGACGGAATTTGTCAAGCAGGAGCAGCTAAGTCTTAAGAAGAA 2280
Db 2336 TATAACTTCTTGTGGGACGGAATTTGTCAAGCAGGAGCAGCTAAGTCTTAAGAAGAA 2395
Qy 2281 GGAGATAATGCACTTCTTAGATACCTGCTGGACAGGATGATCCTAGTATGATCACTCTC 2340
Db 2396 GGAGATAATGCACTTCTTAGATACCTGCTGGACAGGATGATCCTAGTATGATCACTCTC 2455
Qy 2341 TAAAGAACTACAGCCCCAAGTGGGAAGAGTGGACAATAAAATGAGTCAGTGCACAGCTC 2400
Db 2456 TAAAGAACTACAGCCCCAAGTGGGAAGAGTGGACAATAAAATGAGTCAGTGCACAGCTC 2515
Qy 2401 CACCATCTCTAGTCAAGTCAAGAGAAAGACCCCTAAAAATTAAGACAGACAGAGTGAAGA 2460
Db 2516 CACCATCTCTAGTCAAGTCAAGAGAAAGACCCCTAAAAATTAAGACAGACAGAGTGAAGA 2575

Qy	2461	GGGATCTGGAGACTTGGATAAATCTAGATGCTATCTTGGTGATGATCTGACTAGTTCTGACTT	2520
Db	2576	GGGATCTGGAGACTTGGATAAATCTAGATGCTATCTTGGTGATGATCTGACTAGTTCTGACTT	2635
Qy	2521	TTACAATAATTCATATCCTCAAAATGGTAGTCACTGGGGACTAAGCAACAGAGTGTTC	2580
Db	2636	TTACAATAATTCATATCCTCAAAATGGTAGTCACTGGGGACTAAGCAACAGAGTGTTC	2695
Qy	2581	AGGAACATAATTTCTCTGGGTTTGAAAGTTTCACAGTCTGTGCAGTCTATTCTGCTCCTCCATA	2640
Db	2696	AGGAACATAATTTCTCTGGGTTTGAAAGTTTCACAGTCTGTGCAGTCTATTCTGCTCCTCCATA	2755
Qy	2641	TAAACGAGCAGTGTCTCTGGATAGCCCTGTGTTCTTGGTGGCTCAAGTCTCCAGTAAAAA	2700
Db	2756	TAAACGAGCAGTGTCTCTGGATAGCCCTGTGTTCTTGGTGGCTCAAGTCTCCAGTAAAAA	2815
Qy	2701	TATCAGTGTCTTCCCATGTTTACCAAAAGCAACCCATGTGGTGGGAAATCCAGAATGAT	2760
Db	2816	TATCAGTGTCTTCCCATGTTTACCAAAAGCAACCCATGTGGTGGGAAATCCAGAATGAT	2875
Qy	2761	GGATAGTCAGGAAAATTTATGGCTCAAGTATGGGTGGGCCAAACCGAAATGTGACTGTGAC	2820
Db	2876	GGATAGTCAGGAAAATTTATGGCTCAAGTATGGGTGGGCCAAACCGAAATGTGACTGTGAC	2935
Qy	2821	TCAGACTCCTTCTCAGGAGACTGGGGCTTACCAAACTCAAAAGCCCGCAGAAATGGAACC	2880
Db	2936	TCAGACTCCTTCTCAGGAGACTGGGGCTTACCAAACTCAAAAGCCCGCAGAAATGGAACC	2995
Qy	2881	TATGAATTCAAATCCATGGGAAGACACAGGAGGAGATTAATACTCTTTTACCCAGACC	2940
Db	2996	TATGAATTCAAATCCATGGGAAGACACAGGAGGAGATTAATACTCTTTTACCCAGACC	3055
Qy	2941	TGCACGTGGTGGCTCTATTCCACACATTTGCTCTTCGGTCTTAATAGCATACCAAGTGGCAG	3000
Db	3056	TGCACGTGGTGGCTCTATTCCACACATTTGCTCTTCGGTCTTAATAGCATACCAAGTGGCAG	3115
Qy	3001	ACCAGATTGCAACAGCAGCAGATGCTTCAAATAGGCGCTGGTGAATCCCCATGGG	3060
Db	3116	ACCAGATTGCAACAGCAGCAGATGCTTCAAATAGGCGCTGGTGAATCCCCATGGG	3175
Qy	3061	AATGGGGCTTAATCCCTATGGCCAGCAGCAGCATCTAACCAACTGGGTCTTGGGCCCGA	3120
Db	3176	AATGGGGCTTAATCCCTATGGCCAGCAGCAGCATCTAACCAACTGGGTCTTGGGCCCGA	3235
Qy	3121	TGCGATGTTGTCATCGAACAAGTTTCTCATGGCACTCAAATAAGGCTCTTCTTAGGAA	3180
Db	3236	TGCGATGTTGTCATCGAACAAGTTTCTCATGGCACTCAAATAAGGCTCTTCTTAGGAA	3295
Qy	3181	TTCCCTGGATGATCTTGTGGGCCACCTTCCAACTGGAAGCCAGAGTGACGAAAGAGC	3240
Db	3296	TTCCCTGGATGATCTTGTGGGCCACCTTCCAACTGGAAGCCAGAGTGACGAAAGAGC	3355
Qy	3241	ATTATTGGACCAGCTGCACACTTCTCTAGCNAACACAGATCGGACAGGCTTGGAGAAAT	3300
Db	3356	ATTATTGGACCAGCTGCACACTTCTCTAGCNAACACAGATCGGACAGGCTTGGAGAAAT	3415
Qy	3301	TGACAGAGCTTTGGGCAATTCCTGAACTTGCTCAATCAGGACAGAGCATTTAGAGCCCAACA	3360
Db	3416	TGACAGAGCTTTGGGCAATTCCTGAACTTGCTCAATCAGGACAGAGCATTTAGAGCCCAACA	3475
Qy	3361	GGATGCTTTCCAAAGGCCAAGAAGCAGCAGTAAATGATGATCAGAGGCAGGATTTATG	3420
Db	3476	GGATGCTTTCCAAAGGCCAAGAAGCAGCAGTAAATGATGATCAGAGGCAGGATTTATG	3535
Qy	3421	ACAGACATACCCAGCAGAGGGGCTTCCAAATGCAAGGAGGCTTTCATCTTCAGGCAATC	3480
Db	3536	ACAGACATACCCAGCAGAGGGGCTTCCAAATGCAAGGAGGCTTTCATCTTCAGGCAATC	3595
Qy	3481	ACCATCTTTTAACTCTATCATGAATCAGATGAACAGCAGGCAATTTCTCTCCAAAG	3540
Db	3596	ACCATCTTTTAACTCTATCATGAATCAGATGAACAGCAGGCAATTTCTCTCCAAAG	3655
Qy	3541	AATGCAACCCAGCAGCAACATCATGAGACCCCGGACAAACACCCCAAGCAACTTAGAAT	3600

[illegible]

RESULT 2
V99919
ID V99919 standard; CDNA; 4621 BP.
AC V99919;
DT 12-MAY-1999 (first entry)

Db 1498 TGGTCTTGGTCCCAACAGCAGAACATCATGATTTCCCTCGGAATCGTGGCAGCCCAA 1557
QY 1558 GATAGCCTCACAATCAGTTTTCTCTGTTGCAAGTGTGCACCTCTCCCATGGCATCTTCTGG 1617
Db 1558 GATGGCCTCCACCAGTCTCTCTGCTGCTGAGGTGCACACTCACCCATGGACCTTCTGG 1617
QY 1618 CAATACCTGGGNACACAGCTTTTCCACAGAGTCTCTCAGTGGCCCTGCAAGCCATCAGTGA 1677
Db 1618 CAACACAGAGGAGCCACAGCTTTTCTAGCAGTCCCTCAGTGGCTTGAAGCCATCAGTGA 1677
QY 1678 AGGTGCGGGACTTCCCTTTATCTACTCTGTCTATCACCAGGCCCAAAATTTGGATACCTC 1737
Db 1678 AGGCGTGGGACCTCTCTTTATCTCTGTCTCACCAGGCCCAAACTTGGATAATTC 1737
QY 1738 TCCCAATATGAATATTACCAACCAAGTAAGTAAGCAATCAGGATTTCCAAGAGTCCCTCT 1797
Db 1738 TCCCAATATGAATATAAGCCAGCCAACTAAAGTGAAGTGGTCAAGACTCTTAAGAGCCCTT 1797
QY 1798 GGGCTTTATTTGGACCAAAATCCAGTGGAGAGTTCAATGTGTCAAGTTCAAATAGCAGAGA 1857
Db 1798 AGGCTTATCTGTGAACAGAAATCCAGTGGAGAGTTCAAGTGTCAAGTTCAAACAGCAGAGA 1857
QY 1858 TCACCTCAGTGACAAAGAAAGTAGGAGAGCAGTGTTCAGGGGCGCAGAGATCAAAGGG 1917
Db 1858 TCACCCAAAGTGAAAGAAAGAACAGGAGAGCAGTGGGGAGGTGTCAAGAGCGGCCAGGG 1917
QY 1918 TCCCTTTGGAAGCAAGAGTTCATAAAATTAATCTGCACTTACTTACCTGTCTTCTGATGA 1977
Db 1918 ACCTCTGGAAGCAAGAGCCACAGAAATCTGCTGCACTTACTCAGTGTCTTCCCGACGA 1977
QY 1978 CCGGGGTCTATCTCTCTTGACCAACTCCCCCTTAGATTCAAAGTTGTAAGAAATCTTCTGT 2037
Db 1978 CCGAGGCCATCTCTCTTGACCAACTCCCCCTTGATTCAAACTGCAAGACTCTTCCGT 2037
QY 2038 TAGTGTCAACAGCCCTCTGGAGTCTCTCTCTACATCTGGAGGAGTATCCTCTACATC 2097
Db 2038 TAGTGTCAACAGCCCTCTGGAGTCTCTCTCTCAACATCAGGGACAGTGTCTTCCACCTC 2097
QY 2098 CAATATGCATGGGTCACTGTTTACAAGAGAAGCAGCCGATTTTGACAAAGTTGCTGCGAGAA 2157
Db 2098 CAATGTGATGGGTCTCTGTTGCAAGAGAAGAACCCGATTTTGCAAGTGTGTCAGAA 2157
QY 2158 TGGGAATTCACCACTGAGGTAGCCAGATTTACTGCAAGCCACTGGGAAGACACAG 2217
Db 2158 TGGCAACTCCCCAGCGAGGTGCGCAAGATCACTGCAAGGCCACTGGGAAGGACAGAG 2217
QY 2218 CAGTATAACTTCTTGTGGGAGCGAAATGTTGTCAAGCAGGAGCAGTAAAGTCTTAAGAA 2277
Db 2218 CAGCACTGCTTCTGTGGAGAGGGGA---CAACCCAGGCGAGGAGCAGTGAAGTCTTAAGAA 2274
QY 2278 GAAGGAGAATAATGCATCTTCTAGATACCTGCTGGACAGGAGTATCCTAGTATGACAT 2337
Db 2275 GAAGGAGAATAATGCTCTGTTAGATACCTGCTGGACAGGAGTATGCCCAAGTGTGCT 2334
QY 2338 CTCTAAAGAACTACAGCCCAAGTGGAGGAGTGGACAATAAATAGTCAAGTGCACACAG 2397
Db 2335 TGCCAAAGAGCTGCAGCCCGAGCCAGACAGTGGGGACAGTAACTAGTCAAGTGCACCTG 2394
QY 2398 CTCACCACTTCTAGTCAAGTCAAGAGAAGACCCCTAAATTAAGACAGAGCAAGTGA 2457
Db 2395 CTCACCAATCCCAAGCTCTGGCCAAAGAGAAGACCCCAAAATTAAGACAGAGCAAGCA 2454
QY 2458 AGAGGATCTGGAGACTTGGATATCTAGATGCTATTTCTTGGTATCTGACTAGTCTCTGA 2517
Db 2455 GGAGGTATCGGGAGACTTGGATAATCTAGATGCCATTTCTTGGAGATTTGACCATTTCTGA 2514
QY 2518 CTTTACAATAATTTCCATATCTCTCAATAGTGTAGTCACTCTGGGACTAAGCAACAGTGT 2577
Db 2515 CTCTCAACAATCTTA-----CAATGGGGTCAACCCAGGGGCCAAACAGCAGATGTT 2568
QY 2578 TCAAGGAACATAATCTCTGGGTTTGAAGAGTTACAGTGTGTGAGTCTATTCGTCTCTCC 2637
Db 2569 TGCAGGACCGAGTTCTCTGGGTTTGGCAAGTTCACACAGCCTGTGCAAGTCTGTCTCTCTCC 2628

QY 2638 ATATAACCCGAGCAGTGTCTCTGATAGCCCTGTTTCTGTGGCTCAGTCTCTCCAGTAAA 2697
Db 2629 ATATAACCCGAGCGTGTCTCTGATAGCCCTGTTCTGTGGCTCAGTCTCCGCCAGTGAA 2688
QY 2698 AATATACAGTGTCTTCCCAATGTTTACAAAGCAACCCATGTTGGTGGGAATCCAAGAAT 2757
Db 2689 GAATGTAGTGTCTTCCCTGGTTCACAAACAGCCCATACTGGCTGGGAATCCAAGAAT 2748
QY 2758 GATGGATAGTCAGAAATATGCTCAAGTATGGGTGGGCCAAACCGGAATGTGACTGT 2817
Db 2749 GATGGATAGTCAGGAATATACGGTGCACAACTGGGG---CCAAACAGAAATGTTCCTGT 2805
QY 2818 GACTCAGACTCTCTCTCAGGAGACTGGGCTTACCAAACTCAAAAGCCGCGCAAGATGGA 2877
Db 2806 GAATCCGACTTCTCCCCCGAGACTGGGCTTAGCTAATCAAGGGCCAGCAGATGGA 2865
QY 2878 ACCTATGAATTTAAATCTCCATGGGAAGACCAGGAGGAGATTATAACTTCTTTTACCAG 2937
Db 2866 GCCTCTGGCATCAAGTCCCCCTGGGAAGAACTGGAGCGGATTACAGTCCCACTTTACCAG 2925
QY 2938 ACCTGCACTGGGTGGCTCTATTCCCACATTTGCCCTCTTCGGTCTTAATAGCATACAGGTGC 2997
Db 2926 ACCTGCCATGGGGGGCTCTGTGCTTACCTTGGCACTTCTTAATGCACTGCCAGGTGC 2985
QY 2998 GAGACCAGTATT----- 3010
Db 2986 AAGACCATCGTTGCGACGACGAGCAGCAACAGCAGCAACAGCAACAGCAGCA 3045
QY 3010 -----GCAACAGCAGCAGCAGAGTCTTCAAATGAGGCTTGGTGA 3048
Db 3046 ACAGCAGCAGCAACAGCAGCAGCAACAGCAGCAGATGCTTCAAATGAGAACTGGTGA 3105
QY 3049 AATCCCCATGGGAATGGGGCTTAATCCCTATGGCCAGCAGCAGCATCTAACCAACTGGG 3108
Db 3106 GATTTCCATGGGAATGGGAGTCAATCCCTATAGCCAGCAGTCCCGTCTAACCAACAGG 3165
QY 3109 TTCTTGCCCGATGGGATGTTGCTTCAAGCAAGTTTCTCATGGCACTCAAATAGGCC 3168
Db 3166 TTCTTGCCAGAGGAGTGTCTCTATGGAACAAGTCTCTACGGGTCTCAAATAGGCC 3225
QY 3169 TCTTCTTAGGAATTTCCCTGGATGATCTTGTGGGCCACCTTCCAACCTTGAAGGCCAGAG 3228
Db 3226 TCTTCTTAGAAACTCTCTGGATGATCTGTGTGGCCACCTTCTAACGACAGGSGCCAGAG 3285
QY 3229 TGAGGAAGAGCATTAATGGACAGCTGCACACTCTTCTCAGCAACACAGATGGACAGG 3288
Db 3286 TGACGAGAGAGCTCTGCTGGACAGCTGCACACTCTCTGAGCAACACAGATGCCACAGG 3345
QY 3289 CCTGGGAAGAAATTGACAGAGCTTTGGGCACTTCTGAATTTGCAATCAGGAGCAGGCAAT 3348
Db 3346 TCTGGAGGAGATCGACAGGCGCTTGGGAATCTCTGAGTCTGTAATCAGGAGCAAGCTTT 3405
QY 3349 AGAGCCCAACAGAGTCTTTCCAAGGCCAAGAGCAGCAGTAAATGATGGATCAGAAAGGC 3408
Db 3406 GGAGTCCAAACAGAGTGTTTTCCAAGGCCAAGAGCAGCAGTAAATGATGGATCAGAAAGGC 3465
QY 3409 AGATTTATATGGACAGACATACCCAGCAGAGGCGCTCCAAATGCAAGGAGGCTTTTCACT 3468
Db 3466 TGCACATATATGGACAGACATACCCAGCTCAGGCTCCTCCCTTCAAGGAGGCTTTAACCT 3525
QY 3469 TCAGGGAACATACCACTCTTTTAACTCTATGATGAATCAGATGAACAGCAGCAAGCAATTT 3528
Db 3526 TCAGGGAACATCACTGCTTTTAACTCTATGATGGGTGAGATTAGCCAGCAGGACGCTT 3585
QY 3529 TCCTCTCCAAAGGAATGACCCAGCAGCAACATCATGAGACCCCGGACAAACACCCCAA 3588
Db 3586 TCCTCTGCAAGGATGCATCTTAGACCGGCTCTGTCAGACCAAGGACCAACACCCGAA 3645
QY 3589 GCAACTTAGAATGCAGCTTCAGCAGAGGCTGACAGGCGCAGCAGTTTTTGAATCAGAGCG 3648
Db 3646 GCAGCTGAGAATGCAGCTTCAGCAGAGGCTACAGGGCCAGCAGTTTTTAAATCAGAGCGC 3705

Db 3048 AACAGCAGCAGCAACAGCAGCAGATGCTTCAATGAGAACTGGTGAGATTCCTCATGGAA 3107
Qy 3063 TGGGGCTTAATCCCTATGCGCAAGCAGCAGCATCTAACCAACTGGGTTCCTGGCCGATG 3122
Db 3108 TGGGAGTCAATCCNTATAGCCAGCAGTGCAGTCTAACCAACAGGTTCCTGGCCAGAG 3167
Qy 3123 GCATGTTGTCATGGAACAAGTTTCTCATGGCACTCAAAATAGGCTCTCTCTAGGAAT 3182
Db 3168 GCATGCTCTATGGAACAAGTCTCAGGGTCTCAAAATAGGCTCTCTCTAGAACT 3227
Qy 3183 CCCTGGATGATCTGTTGGGCCACCTTCCAACTCGAAGGCGAGAGTGACGAAGAGCAT 3242
Db 3228 CTCCTGGATGATCTGTTGGGCCACCTTCTAACGCGAGAGGGCGAGAGTACGAGAGGCTC 3287
Qy 3243 TATTGGACAGCTGCACACTCTTCTCAGCAACACAGATGCGCAGGCTCGAAGAAATG 3302
Db 3288 TGCTGGACAGCTGCACACTCTTCTGAGCAACACAGATGCGCAGGCTCGAGGAGATCG 3347
3303 ACAGAGCTTTGGGCAATTCCTGAACTTGTCAATCAGGGGACAGCATTAGAGCCCAACAGG 3362
3348 ACAGGCTTGGGAATTCCTGAGCTGTAATCAGGGACAGCTTTGGAGTCCAAACAGG 3407
Qy 3363 ATGCTTTCCAGGCGCAAGCAGCAGTAATGATGATGATGATGATGATGATGATGATGATG 3422
Db 3408 ATGTTTCCAGGCGCAAGCAGCAGTAATGATGATGATGATGATGATGATGATGATGATG 3467
Qy 3423 AGACATACCCAGCAGCAGGCTTCAATGCAAGGAGGCTTCACTTTCAGGACAACTAC 3482
Db 3468 AGACATACCCAGCAGGCTTCCCTTCAAGGAGGCTTTAACTTTCAGGACAGTAC 3527
Qy 3483 CATCTTTAACTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3542
Db 3528 CATCTTTAACTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3587
Qy 3543 TGCACCCAGCAGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3602
Db 3588 TGCATCTAGAGCGGCTCGTGAGACCAAGCAGCAACACCCGAGCAGTGAATGTC 3647
Qy 3603 AGCTTACAGAGGCTGAGGCGCAGCTTTTGAATCAGAGCGCAGCAGCTTGAAT 3662
Db 3648 AGCTTACAGAGGCTGAGGCGCAGCTTTTGAATCAGAGCGGCGCAGCTTGAAT 3707
Qy 3663 TGAATGGAACCCCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3722
Db 3708 TGAATGGAACCCCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3755
3723 AGCAGGTTTCTTAATGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3782
3756 CCCAGGCTTCTTAAATGCCAAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3815
Qy 3783 ACTTCGACACAGAGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 3842
Db 3816 ACCTCGACAGAGGATGGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 3847
3843 AGCAGCAGCAGCAACAGCAACAGCAACAGCAGCAACAGCAGCAACAGCAGCAGCAGCAG 3902
Db 3847 -----GTCAACACACAGCCTCAGCCTCAGCCTCAGCCTCAGCCTCAGCCTCAGCCT 3875
Qy 3903 CACCTCTTAATGTGACTGCTTCCCGCAGCAGTGGATGGGCTTTTGGCAGGACCCCAATGC 3962
Db 3876 CACCTCCCAAGCTCACCGCTCCCGCAGCAGTGGATGGGCTTTTGGCAGGATGC 3935
Qy 3963 CACAGCTCTCCGCAACAGTTCATATCAACCAATTTGGAATGGGACCAACACAG 4022
Db 3936 CGCAGCCCTCCCAACAGTTCATATCAACCAATTTGGAATGGGACCAACACAG 3995
Qy 4023 -ATCCAGCTTTGGTCGAGTGTCTAGTCTCCCAATGATGATGATGATGATGATGATGATG 4081
Db 3996 TAGCCAGCTTTGGTCGAGGCTCGAGTCTCCAGTGAATGATGATGATGATGATGATGATG 4055
Qy 4082 CCCTTCCAGATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4141
Db 4056, COTTCAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4115

Qy 4142 ATGAAGGCTGGCCATCAGGAATTTGGCCAGGAACAGCTCCTTTTCCAGCAGCAGTTT 4201
Db 4116 ATGAAGGCTGGCCATCAGGAATTTGGCCAGGAACAGCTCCTTTTCCAGCAGCAGTTT 4175
Qy 4202 GCCCACCAGGGAATTCCTGAGTGTATAGTGTGACATGAATGACAGCAGGTTGGTAC 4261
Db 4176 GCTCCACAGGGAATTCCTGAGTGTATAGTGTGACATGAATGACAGCAGGTTGGTAC 4235
Qy 4262 ATGGACAGATGAATGAACCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4321
Db 4236 TTGGACAGATGACATGACCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4295
Qy 4322 AAATACGCTGACATCTCTGACAGGACCTTTTAAAGGAACCACTGTACAAATGACAT 4381
Db 4296 AAATACGCTGACATCTCTGACAGGACCTTTTAAAGGAACCACTGTACAAATGACAT 4348
Qy 4382 GCATAGGATTTATGGAAGGAATCATTTTCCAGGACATCATTTTGAAGAAAGGAC 4439
Db 4349 GGATCATCAGGACCTGGCGGAGTCTATGCTTAAGCATCAGCTTGAAGCAAGGCC 4406

RESULT 4

V03517

ID V03517 standard; cDNA; 6156 BP.

AC V03517;

DT 20-JUL-1998 (first entry)

DE Human transcriptional intermediary factor-2 (TIF2) cDNA.

KW Transcriptional intermediary factor; TIF2; human; drug screening;

KW assay; nuclear receptor; ds.

OS Homo sapiens.

PH Key

FT Location/Qualifiers

CDS 163..4557

/*tag= a

PW09802455-A2.

PN 22-JAN-1998.

PF 11-JUL-1997; US12100.

PR 12-JUL-1996; US-021247.

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

PA (CNRM) CENT NAT RECH SCI.

PA (INRM) INST NAT SANTE & RECH MEDICALE.

PA (UYPA) UNIV PASTEUR LOUIS.

PI Chambon P, Gronemeyer H, Lutz Y, Voegel J;

P-PSDB: W42632.

DR New isolated transcriptional intermediary factor-2 - useful in

PT assays for drugs which are capable of enhancing or inhibiting

PT nuclear receptor-mediated pathways

PS Claim 1; Fig 1A-B; 119pp; English.

CC This cDNA clone codes for a novel nuclear receptor transcriptional

CC mediator (see W42632) designated transcriptional intermediary

CC factor-2 (TIF2). TIF2 interacts directly with the ligand binding

CC domains (LBD) of several nuclear receptors (NR) in an agonist- and

CC AF2-integrity-dependent manner in vitro and in vivo, harbours an

CC autonomous AF, relieves NR autoquenching, and enhances the

CC activity of NR AF2s when overexpressed in mammalian cells.

CC Screening of a human placenta cDNA expression library with an

CC estradiol-bound 32P-labelled estrogen receptor probe yielded a cDNA

CC that encoded a protein fragment (TIF2.1) that interacted on

CC Far-Western blots with different 32P-labelled NR-LBDs ER, RAR, RXR

CC in an agonist-dependent manner. The TIF2 coding sequence was

CC obtained upon rescanning with a TIF2.1 cDNA probe. TIF2 cDNA has

CC been deposited as ARCC 97612. Polynucleotides encoding full-length

CC TIF2 protein or encoding TIF2 polypeptide fragments comprising

CC amino acids 624-869, 624-1131, 1010-1131 or 1288-1464 are claimed.

CC Recombinant methods for making TIF2 polypeptides are provided, as

CC are screening methods for identifying agonists and antagonists of

CC nuclear receptor AF-2 function, TIF2 AD1 activity and TIF2 AD2

CC activity. The products are useful in assays for identifying drugs

CC capable of enhancing or inhibiting NR-mediated pathways. They can

CC also be used for detection and localisation.







CC Sequence 6156 BP; 1799 A; 1490 C; 1406 G; 1461 T;

SQ

[illegible]

[illegible]

This Page Blank (uspto)

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

on: Fri Sep 17 21:02:33 1999; MasPar time 48.85 seconds
616.024 Million cell updates/sec

```

>US-09-041-994-2
Description:
Perfect Score: 9849
Sequence: 1 MSGGENLDPLASDSRRKL.....MNNPMPMSGMPGPDQKYC 1415

Scoring table:
PAM 150
Gap 11

```

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a-geneSeq35

1:part12	2:part2	3:part3	4:part4	5:part5	6:part6	7:part7
8:part8	9:part9	10:part10	11:part11	12:part12	13:part13	
14:part14	15:part15	16:part16	17:part17	18:part18		
19:part19	20:part20	21:part21	22:part22	23:part23		
24:part24	25:part25	26:part26	27:part27	28:part28		
29:part29	30:part30	31:part31	32:part32	33:part33		
34:part34	35:part35	36:part36	37:part37	38:part38		
39:part39						

Statistics: Mean 39.485; Variance 214.637; scale 0.184

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	3346	34.0	1464	30	W42632	2.04e-248
2	914	9.3	147	34	Human transcriptional	1.24e-57
3	643	6.5	1061	24	Clone AJ1.1 protein s	6.23e-37
4	232	2.4	846	37	Human steroid recepto	7.96e-07
5	232	2.4	848	14	Human CLOCK protein.	7.96e-07
6	232	2.4	848	23	Human Ah receptor pro	7.96e-07
7	235	2.4	855	37	Human Ah-receptor.	4.96e-07
8	225	2.3	371	38	Mouse CLOCK protein.	2.39e-06
9	231	2.3	805	14	Epitope tagged TBP pr	9.32e-07
10	213	2.2	1313	33	Murine Ah receptor pr	1.56e-05
11	202	2.1	539	30	Spinoecerebellar ataxi	8.55e-05
12	208	2.1	594	34	Yeast transcriptional	2.90e-05
13	210	2.1	800	14	Mouse neuronal PAS do	2.48e-05
14	202	2.1	870	27	Transcription factor	8.55e-05
15	203	2.1	914	25	Human endothelial PAS	7.33e-05
16	203	2.1	1312	30	Spinoecerebellar ataxi	7.33e-05
17	203	2.1	1312	30	Human ataxin-2.	7.33e-05

aspects of circadian rhythm, including intrinsic circadian period and the persistence of circadian rhythmicity. The sequence was deduced from the isolated clock gene (see V61450). Mouse CLOCK (see W79529) is also claimed. CLOCK is a member of the bHLH-PAS domain family of proteins, and thus likely interacts directly with DNA. It has Gln-, Pro- and Ser-rich C-terminal regions characteristic of activation domain transcription factors. The invention provides isolated and purified CLOCK polypeptides, polynucleotides, vectors and host cells. The polynucleotides or polypeptides can be used to treat disorders of altered or disrupted circadian rhythms e.g. jet-lag, seasonal affective disorder, sleep-wake cycle disorders such as mood state, stress, neurological disorders, to regulate diet and food intake especially for diabetes, to treat cardiovascular, respiratory, liver or endocrine disorders, and for diagnosis and treatment of abnormal cell division such as cancer.

SQ Sequence 846 AA;

Query Match 2.4%; Score 232; DB 37; Length 846;
Best Local Similarity 27.7%; Pred. No. 7.96e-07;
Matches 46; Conservative 48; Mismatches 64; Indels 8; Gaps 7;

39 rnksekrrdgnvlik-elgmlpnarkmdkstvlgksidflrkxhei-taqdsasei 96
QY 37 RREQESKYEELAEELISANLSDIDNFV-KPKCAILKETVRIQIRKQEGKTIENDDV 95
Db 97 -xqg--wk-ptflsneeftqlmealdgfflaimgdgsliyvsesvtsllehlpsdlvdq 152
QY 96 QRADVSTCGQVIDKDSLGSLGALLQALDGLFVVNREANIVFVSENVTQYLOKQEDLVNT 155
Db 153 slnfipegehevskilthlesdsltpeylksnqlfchml 198
QY 156 SVYNILHEEDRKDFLKNPKSTVNGSVTNEPQKQKSH-TFNCRLM 200

RESULT 5

ID R80551 standard; Protein: 848 AA.
AC R80551;
DT 29-NOV-1995 (first entry)
DE Human Ah receptor protein.
KW Mouse Ah receptor; drug; carcinogen; toxic agent; probe; human; bioassay;
KW C57BL/6J mouse liver; environmental pollutant; recombinant organism;
KW biomonitor.
OS Homo sapiens.
PN US5378822-A.
PD 03-JAN-1995.
PF 08-APR-1993; 045806.
PR 08-APR-1993; US-045806.
(NOUN) UNIV NORTHWESTERN & FOUND.
(WISC) WISCONSIN ALUMNI RES FOUND.
Bradfield CA, Dolwick KM, Poland A;
WPI: 95-051315/07.
DR N-PSDB; Q99602.
PT Nucleic acid sequences encoding murine and human Ah receptors -
PT used in competitive binding assays to detect environmental
PT pollutants
PS Claim 3; Column 41-48; 24pp; English.
CC The amino acid sequence of the human Ah receptor protein. The Ah receptor
CC is a soluble protein which mediates an individual's response to a variety
CC of drugs, carcinogens and toxic agents. The human gene was isolated from
CC a cDNA library derived from hepatoma cell line HepG2 mRNA using as a
CC probe a 1.4 kb EcoRI fragment from the murine Ah receptor gene (Q99601).
CC This screen isolated a clone designated 91A which contained a 4.47 kb
CC insert. The insert started with a continuous open reading frame (ORF)
CC encoding 732 amino acids prior to an in-frame termination codon. A 920 bp
CC BamHI fragment of clone 91A was used as a probe to isolate a clone (hul4)
CC which overlapped by 1.56 kb with 91A. Sequence analysis of hul4
CC the ORF of 91A by an additional 116 amino acids to an initiation
CC methionine. Clone 91A also contains 2.27 kb of 3' untranslated region
CC (UTR). To complete the 3' UTR, primers Q99608-9 were used in a rapid
CC amplification of cDNA ends (RACE) method. The Ah receptors are useful in
CC bioassays to detect environmental pollutants. The genes can be used to
CC generate recombinant organisms useful as biomonitors for environmental
CC pollutants.

SQ Sequence 848 AA;
Query Match 2.4%; Score 232; DB 14; Length 848;
Best Local Similarity 27.5%; Pred. No. 7.96e-07;
Matches 42; Conservative 52; Mismatches 50; Indels 9; Gaps 7;
Db 27 paegiksnpskrrhrldnteldrlasllp--fpq-dvin-kldklsvrlsvlyraksf 82
QY 25 PGOGLTCSGERRRRQESKYEELAEELISANLSDIDNFNPKCAILKETVRIQIRK- 83
Db 83 fdvalkspternggdncraanfrinlge-geflqalngfvlvvttdalvfyasat 141
QY 84 -EQG-KTISNDDVQKADVSTG-QGVIDKDSLGSLGALLQALDGLFVVNREANIVFVSEN 140
Db 142 lqdylgfsgsdvthgsvyellhtedraefqrql 174
QY 141 VTQYLOKQEDLVNTSVYNILHEEDRKDFLKNL 173

RESULT 6

ID W25668 standard; Protein: 848 AA.
AC W25668;
DT 05-NOV-1997 (first entry)
DE Human Ah-receptor.
KW Human; Ah receptor; yeast; nuclear translocator; reporter gene;
KW dioxin responsive element; environmental pollutant.
OS Homo sapiens.
PN Location/Qualifiers
FT key 13..41
FT region /note= "Basic region"
FT domain 43..84
FT domain /note= "Helix-loop-helix domain"
FT domain 113..400
FT domain /note= "PAAS domain"
PN US5650283-A.
PD 22-JUL-1997.
PF 08-APR-1993; 045806.
PR 29-DEC-1994; US-366051.
PR 08-APR-1993; US-045806.
PA (NOUN) UNIV NORTHWESTERN.
PI Bradfield CA, Carver LA, Dolwick KM;
DR WPI: 97-384667/35.
DR N-PSDB; T85436.
PT New genetically engineered yeast and mammalian cells - contain the
PT Ah receptor gene and a reporter gene, used for detecting Ah receptor
PT agonists, particularly environmental pollutants
PS Claim 1; Fig 6; 47pp; English.
CC This sequence represents the human Ah receptor molecule. The cDNA
CC encoding this protein may be used to transform a viable yeast cell,
CC which may also be transformed with a plasmid expressing an Ah receptor
CC nuclear translocator, a dioxin responsive element and a reporter gene.
CC The reporter gene detects the activation of the Ah receptor upon the
CC binding of agonists to the Ah receptor. The transformed cells can be
CC used to detect agonists to the Ah receptor in environmental samples such
CC as air, water and soil or in tissue samples. They can be used for
CC detecting and monitoring environmental pollutants such as dioxin.

SQ Sequence 848 AA;

Query Match 2.4%; Score 232; DB 23; Length 848;
Best Local Similarity 27.5%; Pred. No. 7.96e-07;
Matches 42; Conservative 52; Mismatches 50; Indels 9; Gaps 7;

Db 27 paegiksnpskrrhrldnteldrlasllp--fpq-dvin-kldklsvrlsvlyraksf 82
QY 25 PGOGLTCSGERRRRQESKYEELAEELISANLSDIDNFNPKCAILKETVRIQIRK- 83
Db 83 fdvalkspternggdncraanfrinlge-geflqalngfvlvvttdalvfyasat 141
QY 84 -EQG-KTISNDDVQKADVSTG-QGVIDKDSLGSLGALLQALDGLFVVNREANIVFVSEN 140
Db 142 lqdylgfsgsdvthgsvyellhtedraefqrql 174
QY 141 VTQYLOKQEDLVNTSVYNILHEEDRKDFLKNL 173

WPI; 95-051315/07.
N-PsDB; Q99601.
Nucleic acid sequences encoding murine and human Ah receptors -
used in competitive binding assays to detect environmental
pollutants

Claim 1; Column 31-36; 24pp; English.

The amino acid sequence of the murine Ah receptor protein. The Ah
receptor is a soluble protein which mediates an individuals response to a
variety of drugs, carcinogens and toxic agents. The gene was isolated
from a mouse genomic DNA library using the probes Q99603-5. These probes
were determined from the N-terminal sequence of a purified Ah receptor,
which was purified from C57Bl/6J mouse liver. The screen isolated a
clone designated cAh1 containing the full length cDNA. The 1.4 kb
fragment of the murine clone was used to obtain clones covering the
coding sequence of the human AH receptor (Q99602). The Ah receptors are
used in bioassays to detect environmental pollutants. The genes can be
used to generate recombinant organisms useful as biomonitors for
environmental pollutants.
Sequence 805 AA;

Query Match 2.3%; Score 231; DB 14; Length 805;
Best Local Similarity 30.3%; Pred.No. 9.32e-07;
Matches 46; Conservative 46; Mismatches 52; Indels 8; Gaps 6;

Ddb 26 paegikspksrhrdrInteldriasllp--fpq-dvin-kidklsvlrivstvlraksf 81
::: : |||:: : ::||:: : ::||:: : ::||:: : ::||:: : ::||:: : : |
QOY 25 PGOGLTCSGERRRREQESKYEELAEILISANLSDIDNFNVKPKCAIKETVRIQRKE 84
::: : |||:: : ::||:: : ::||:: : ::||:: : ::||:: : ::||:: : : |
Ddb 82 fdvalkstpadrngggdqcrsq-lrdwqlqgegfllqalgngflvvtadalyfassti 140
::: : |||:: : ::||:: : ::||:: : ::||:: : ::||:: : ::||:: : : |
QOY 85 OCKTI-SNDDDVQKADSVSTGGVIDKDLSL-GPLLQLALGDGFLFVNREANIVFSENV 141
::: : |||:: : ::||:: : ::||:: : ::||:: : ::||:: : ::||:: : : |
Ddb 141 qdyiqfgsgdvihqsyyelihtedraefqrql 172
::: : |||:: : ::||:: : ::||:: : ::||:: : ::||:: : ::||:: : : |
QOY 142 TQYLQKGEDLVNTSVYNILHEEDRKDFLNKL 173
::: : |||:: : ::||:: : ::||:: : ::||:: : ::||:: : ::||:: : : |

RESULT 10
IID W60213 standard; Protein; 1313 AA.
AC W60213;
DE 02-OCT-1998 (first entry)
SITE Spinocerebellar ataxia type 2 (SCA2) disease associated protein.
KW Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KK CAG repeat; neurodegenerative disease.
QQ Homo sapiens.
FF Key Location/Qualifiers
FT Region 166..188
/note= "Glutamine rich region; this region is increased in SCA2 patients"

WO9818920-A1.
PD 07-MAY-1998.
PF 30-OCT-1997; J03946.
PPR 30-OCT-1996; JP-304059.
PAPA (SRLS-) SRL INC.
PI Sanpei K, Tsuji S;
PI WPI: 98-272215/24.
DR N-PsDB: V30270.
PPT Nucleic acid fragments associated with spinocerebellar ataxia type 2
PPT contain increased number of CAG repeat region compared to normal
PPT gene
PPT Claim 1; Pages 13-22; 38pp; Japanese.
CCC This is the protein sequence of a gene causative of spinocerebellar
CCC ataxia type 2 (SCA2), a neurodegenerative disease. The gene associated
CCC with SCA2, has a tri-nucleotide (CAG) repeat region which in the
CCC expression product produces a polyglutamine sequence from Gln-166 to
CCC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
CCC patients this number is increased to 35-100. Peptides encoded by nucleic
CCC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
CCC gene, antibodies recognising the peptides and antisense nucleic acids
CCC hybridising with the nucleic acid fragments can be used for the
CCC investigation and diagnosis of SCA2. They can also be used for the
CCC treatment of SCA2 by antisense therapy or gene therapy.
SQ .sequence 1313 AA;


```
CC suffering from spinocerebellar ataxia (SCA), with MAB 1C2 isolated 6
CC new sequences (T78906-T78911) encoding polyglutamine repeats. MAB 1C2
CC also isolated the complete SCA2 gene in clone DAN1 which appears to
CC contain 2 open reading frames (ORF), the second of which may be generated
CC by an frameshift slippage or by an alternative splicing event. This
CC sequence represents the first ORF encoded by the SCA2 gene. The protein
CC contains a 22 amino acid polyglutamine region. Normal SCA2 proteins
CC contain 17-29 glutamines in the repeat whereas the mutant sequence
CC from patients with SCA contains at least 30, preferably 37-50 repeats.
CC MAB 1C2, active fragment of it or nucleic acids encoding it are
CC specifically used to treat Huntington's disease, SCA types 1-5 or 7,
CC X-linked spino-bulbar muscular atrophy (Kennedy disease),
CC dentatorubral-pallidoluysal atrophy, dominant autosomal spinocerebellar
CC ataxia, familial spastic paraplegia, bipolar affective disorder, manic
CC depressive psychoses and schizophrenia.
CC Sequence 914 AA;

Query Match      2.1%; Score 203; DB 25; Length 914;
Best Local Similarity 53.4%; Pred.No. 7.33e+05;
Matches 31; Conservative 13; Mismatches 13; Indels 1; Gaps 1;

Db 78 pLtkspkqgggqqgggqqgggqqgggqpaaanvr-kpgggslaspaapaps 134
   : | ||||| ||||| ||||| ||||| ||||| ||||| ||||| : : :
QY 1240 AMMQQQQQQQQQQQQQQQQQQQQQQQQQQQQTATSPFPNTASMSDGLAGTRMPQAPP 1297

Search completed: Fri Sep 17 21:06:34 1999
Job time : 241 secs.
```

MILES

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

on: Fri Sep 17 21:16:43 1999; MasPar time 16.36 Seconds
878.195 Million cell updates/sec

Tabular output not generated.

Title: >US-09-041-994-2
Description: (1-1415) from US09041994.pep
Perfect Score: 9849
Sequence: 1 MSGLGENDPLASDSRKRL.....MNNPMPMSGMPGPDQKYC 1415

Scoring table: PAM 150
Gap 11

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PC9_COMB 4:backfiles1

Statistics: Mean 36.592; Variance 199.242; scale 0.184

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	232	2.4	848	1 US-08-366-	Sequence 4, Applicatio	1.87e-07
2	232	2.4	848	1 US-08-045-	Sequence 4, Applicatio	1.87e-07
3	235	2.4	855	2 US-08-816-	Sequence 2, Applicatio	1.16e-07
4	231	2.3	805	1 US-08-366-	Sequence 2, Applicatio	2.19e-07
5	231	2.3	805	1 US-08-045-	Sequence 2, Applicatio	2.19e-07
6	210	2.1	360	2 US-08-531-	Sequence 2, Applicatio	6.09e-06
7	209	2.1	594	2 US-08-785-	Sequence 6, Applicatio	7.13e-06
8	202	2.1	870	1 US-08-785-	Sequence 4, Applicatio	2.13e-05
9	210	2.1	2414	3 PCT-US95-0	Sequence 2, Applicatio	6.09e-06
10	210	2.1	2414	1 US-08-227-	Sequence 2, Applicatio	6.09e-06
11	205	2.1	2441	2 US-08-194-	Sequence 2, Applicatio	1.34e-05
12	204	2.1	3144	1 US-08-246-	Sequence 6, Applicatio	1.56e-05
13	204	2.1	3144	1 US-08-453-	Sequence 6, Applicatio	1.56e-05
14	204	2.1	3144	2 US-08-457-	Sequence 42, Applicati	1.56e-05
15	197	2.0	303	2 US-08-185-	Sequence 5, Applicatio	4.65e-05
16	199	2.0	590	2 US-08-785-	Sequence 5, Applicatio	3.41e-05
17	197	2.0	737	2 US-08-185-	Sequence 4, Applicatio	4.65e-05
18	197	2.0	737	2 US-08-185-	Sequence 2, Applicatio	4.65e-05
19	193	2.0	788	2 US-08-918-	Sequence 4, Applicatio	8.66e-05
20	194	2.0	816	2 US-08-267-	Sequence 9, Applicatio	7.41e-05
21	200	2.0	875	1 US-08-785-	Sequence 5, Applicatio	2.92e-05
22	189	1.9	373	3 PCT-US96-1	Sequence 3, Applicatio	1.61e-04
23	189	1.9	373	2 US-08-480-	Sequence 3, Applicatio	1.61e-04

OY 156 SVINILHEEDRKDFLKNPKSTVNGSVWTEPNQOKSH-TFNCRL 200

RESULT 4
ID US-08-366-051B-2 STANDARD; PRT: 805 AA.

XX xxxxxx

Sequence 2, Application US/08366051B

Sequence 2, Application US/08366051B

Patent No. 5650283

GENERAL INFORMATION:

APPLICANT: Bradfield, Christopher A.

APPLICANT: Dolwick, Kristin M.

APPLICANT: Carver, Lucy A.

TITLE OF INVENTION: Ah Receptor cDNAs and Genetically

TITLE OF INVENTION: Engineered Cells for Detecting Agonists to the Ah

TITLE OF INVENTION: Receptor

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut

STREET: 100 South Wacker Drive, Suite 960

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606-4002

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/366,051B

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Tilton, Timothy L.

REGISTRATION NUMBER: 16,976

REFERENCE/DOCKET NUMBER: NU-9207-CIP

TELEPHONE: (312)-456-8000

TELEFAX: (312)-456-7776

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 805 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 805 AA; 90351 MW; 3331614 CN;

Query Match 2.3%; Score 231; DB 1; Length 805;

Best Local Similarity 30.3%; Pred. No. 2.19e-07;

Matches 46; Conservative 46; Mismatches 52; Indels 8; Gaps 6;

Db 26 PAEGIKSNPSRRHRDLNTELDRLASLLP--FPQ-DVIN-KLDKLSVLRLSVTLRAKSF 81

OY 25 PQGLTSGERRRREQESKYIEELAEISANLSDIDNFNVPDRCAILKVTROIQIKE 84

Db 82 FDVALKSTPADRNGQDQCRAQ-IRDMQDLOGEFLQALNGFVLVVTADALVFYASSTI 140

OY 85 QGKTI-SNDDVDQKADVSSTGGVIDKDSL--GPLLLQALDGLFLVYNREANIVFVSENV 141

Db 141 QDYLGFGQSDVIHOSVYELIHTEHEDRAEFQROL 172

OY 142 TQYLQYKQEDLVNTSVYNILHEEDRKDFLKNL 173

RESULT 5
ID US-08-045-806-2 STANDARD; PRT: 805 AA.

XX xxxxxx

Sequence 2, Application US/08045806

Sequence 2, Application US/08045806

Patent No. 5378822

GENERAL INFORMATION:

APPLICANT: Bradfield, Christopher Alan

APPLICANT: Dolwick, Kristin Marie

APPLICANT: Poland, Alan

TITLE OF INVENTION: Ah Receptor cDNA and Method of

TITLE OF INVENTION: Determining Human Risks To Environmental Pollutants

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut

STREET: 100 South Wacker Drive, Suite 960

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606-4002

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/045,806

FILING DATE: 19930408

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Fentress, Susan B.

REGISTRATION NUMBER: 31,327

REFERENCE/DOCKET NUMBER: NU-9207

TELEPHONE: (312)-456-8000

TELEFAX: (312)-456-7776

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 805 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 805 AA; 90351 MW; 3331614 CN;

Query Match 2.3%; Score 231; DB 1; Length 805;

Best Local Similarity 30.3%; Pred. No. 2.19e-07;

Matches 46; Conservative 46; Mismatches 52; Indels 8; Gaps 6;

Db 26 PAEGIKSNPSRRHRDLNTELDRLASLLP--FPQ-DVIN-KLDKLSVLRLSVTLRAKSF 81

OY 25 PQGLTSGERRRREQESKYIEELAEISANLSDIDNFNVPDRCAILKVTROIQIKE 84

Db 82 FDVALKSTPADRNGQDQCRAQ-IRDMQDLOGEFLQALNGFVLVVTADALVFYASSTI 140

OY 85 QGKTI-SNDDVDQKADVSSTGGVIDKDSL--GPLLLQALDGLFLVYNREANIVFVSENV 141

Db 141 QDYLGFGQSDVIHOSVYELIHTEHEDRAEFQROL 172

OY 142 TQYLQYKQEDLVNTSVYNILHEEDRKDFLKNL 173

RESULT 6
ID US-08-531-927B-2 STANDARD; PRT: 360 AA.

XX xxxxxx

Sequence 2, Application US/08531927B

CC	STREET: 268 BUSH STREET, SUITE 3200
CC	CITY: SAN FRANCISCO
CC	STATE: CALIFORNIA
CC	COUNTRY: USA
CC	ZIP: 94104
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.30
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: US/08/785,310A
CC	FILING DATE: 21-JAN-1997
CC	CLASSIFICATION: 536
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: OSMAN, RICHARD A.
CC	REGISTRATION NUMBER: 36,627
CC	REFERENCE/DOCKET NUMBER: UTSD:1226
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: (415) 343-4341
CC	TELEFAX: (415) 343-4342
CC	INFORMATION FOR SEQ ID NO: 6:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 594 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: peptide
CC	SEQUENCE 594 AA; 63736 MW; 1871071 CN;
CC	Query Match 2.1%; Score 209; DB 2; Length 594;
CC	Best Local Similarity 28.5%; Pred. No. 7.13e-06;
CC	Matches 35; Conservative 35; Mismatches 50; Indels 3; Gaps
Db	123 RRGVVALVSE-VFEQH-LGGHILQSLDGFVALNQEGKFLYISVTSYIYGLSQVLELTS 180
Qy	96 QKADVSSTGGGVGDKDGLGPLLQALDGLFVNVNREANIVFVSSENVTQYLYKQEDLVNT 155
Db	181 SVFDYIHPGDHSEVLEOLGLRAAS-ICPPTPPSVSSSSSSSSSSSLVDVTPLEIASPTEASP 239
Qy	156 SVYILHEEDRKDKFLNPKXSTVNGVSWTNEPQRKSHFTNCRMLMKTPHDILEDINASP 215
Db	240 AFR 242
Qy	216 EMR 218
RESULT 8	
ID	US-08-785-241-4 STANDARD: - PRT: 870 AA.
XX	xxxxxx
AC	
XX	
DT	
XX	
DE	
XX	
XX	
CC	Sequence 4, Application US/08785241
CC	Sequence 4, Application US/08785241
CC	Patent No. 5695963
CC	GENERAL INFORMATION:
CC	APPLICANT: McKnight, Steven L.
CC	APPLICANT: Russell, David W.
CC	APPLICANT: Tian, Hui
CC	TITLE OF INVENTION: Endothelial PAS Domain Protein
CC	NUMBER OF SEQUENCES: 7
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
CC	STREET: 268 BUSH STREET, SUITE 3200
CC	CITY: SAN FRANCISCO
CC	STATE: CALIFORNIA
CC	COUNTRY: USA
CC	ZIP: 94104
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk

XX Sequence 2, Application US/08194468
 DE
 XX
 CC Sequence 2, Application US/08194468
 CC Patent No. 5750336
 CC
 CC GENERAL INFORMATION:
 CC APPLICANT: Montminy, Marc R.
 CC TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
 CC TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
 CC TITLE OF INVENTION: RESPONSIVE GENES
 CC NUMBER OF SEQUENCES: 3
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Pretty Schroeder, Brueggemann & Clark
 CC STREET: 444 South Flower Street, Suite 2000
 CC City: Los Angeles
 CC STATE: California
 CC COUNTRY: USA
 CC ZIP: 90071
 CC
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC Compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patentin Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/194,468
 CC FILING DATE: 10-FEB-1994
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Reiter, Stephen E.
 CC REGISTRATION NUMBER: 31,192
 CC REFERENCE/DOCKET NUMBER: P41 9672
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (619)-546-4737
 CC

CC	SEQUENCE	2441 AA; 265473 MW; 30223014 CN;
CC	Query Match	2.1%; Score No. 1; Length 2441;
CC	Best Local Similarity	29.4%; Pred. No. 1.34e-05;
CC	Matches	65; Conservative 45; Mismatches 97; Indels 14; Gaps 12;
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	2441 amino acids
CC	TYPE:	amino acid
CC	TOPOLOGY:	linear
CC	MOLECULE TYPE:	protein
CC	SEQUENCE	2441 AA; 265473 MW; 30223014 CN;
Db	2116 MQPQGLQSQGMPQPGHGHQPSL-QNLNMQAGVPRGVPVPPQAMGGLNPQGALNI	2174
Qy	1160 MRPTNTPKQLRMQLQRQQQLNQSRQAELKKNENPTAGAAVMRPMWQPQGFLLA	1219
Db	2175 -MNPGNPNMTNN-PQYR-EMVRRQLQHQQQQQQQQQQQQQQNSASLAGGMAGHSQF	2231
Qy	1220 QMVARSRELLSHFRQQRVAMMMQSQQQQQQQQQQQQQQQQTQAFSPPPNV-	1278
Db	2232 QQPQGGGYA-PAMQQRMQHLPIQGS-SMGMAAP-MGQLG--Q-MGPQLGADSTP	2284
Qy	1279 ASFSMDGLLAGTPMPQAP-PQFPYPYNGMGQPDPAFGRVSSPNNAMSSRWGFSQNP	1337
Db	2285 NIQAALOORLIQQQMKQ-QIGSPGPNPMSPOOHLMSGP	2324
Qy	1338 MMQHPQAASIYSSEKMGWPSGLARNSSFSQQFAHQGNP	1378
RESULT	12	
ID	US-08-246-982A-6	STANDARD; PRT; 3144 AA.
XX	xxxxxx	
AC		
DT		
XX		
DE	Sequence 6, Application US/08246982A	
CC	Sequence 6, Application US/08246982A	


```
Matches      30; Conservative    24; Mismatches   29; Indels     0; Gaps       0;
Db    213 PQOANOLKSNSASVSSQYNTLPKLGDTKSLHRVPMTQQOHPLPTSHOVQQQHQLQHQQQ 272
        :::|| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Qy    1187 SRQAELKKMENPTAGGAAYVRPMQSQPQGFLNAQMVAQRSELLSHHFRRQRYAMMMMQQ 1246

Db    273 QOOQHHRHHHQHQQHQQHQQHQM 295
        ||||||::|:|||::|:|||::|::|
Qy    1247 QQOQQOQQOQQOQQOQQOQQOQTQ 1269
```

Search completed: Fri Sep 17 21:17:45 1999
Job time : 62 secs.

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 19:18:26 ; Search time 1821.02 Seconds
(without alignments)
7852.012 Million cell updates/sec

Title: US-09-041-994-1
Perfect score: 4496
Sequence: 1 GCTGATGGTGACTCAGAG.....CATTTGACGAGGAAATCTAG 4496

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

abase : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_st.*
- 14: gb_sts.*
- 15: gb_sy.*
- 16: gb_un.*
- 17: gb_vi.*
- 18: em_fun.*
- 19: em_htg.*
- 20: em_hum1.*
- 21: em_hum2.*
- 22: em_in.*
- 23: em_om.*
- 24: em_or.*
- 25: em_ov.*
- 26: em_pat.*
- 27: em_ph.*
- 28: em_pi.*
- 29: em_ro.*
- 30: em_sts.*
- 31: em_sy.*
- 32: em_un.*
- 33: em_vi.*
- 34: gb_htg1.*
- 35: gb_htg2.*
- 36: gb_in1.*
- 37: gb_in2.*
- 38: gb_ba1.*
- 39: gb_ba2.*
- 40: em_hum3.*
- 41: em_hum4.*
- 42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4468.2	99.4	4495	11	AF010227 Homo sapi

ALIGNMENTS

RESULT 1

LOCUS	AF010227	4495 bp	mRNA	PRI	13-AUG-1997
DEFINITION	Homo sapiens receptor-associated coactivator 3 (RAC3) mRNA,				
ACCESSION	AF010227				
NID	92318005				
VERSION	AF010227.1	GI:2318005			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Li, H., Gomes, P.J., and Chen, J.D.				
TITLE	RAC3, a steroid/nuclear receptor-associated coactivator that is related to SRC-1 and TIF2				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 94 (16), 8479-8484 (1997)				
MEDLINE	97385128				
REFERENCE	2 (bases 1 to 4495)				
AUTHORS	Chen, J.D., Li, H., and Gomes, P.J.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-JUN-1997) Pharmacology and Molecular Toxicology, University of Massachusetts Medical School, 55 Lake Avenue North,				

Worcester, MA 01655-0126, USA

FEATURES

Location/Qualifiers

source

1. .4495

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_line="HeLa"

1. .4495

gene

/gene="RAC3"

86. .4339

CDS

/gene="RAC3"

/note="transcriptional coactivator with intrinsic->histone acetyltransferase activity; member of the steroid/nuclear receptor-associated coactivator family which includes steroid receptor coactivator 1 (SRC-1), transcriptional intermediate factor 2 (TIF2), and receptor associated coactivator 3; similar to the mouse p300/CBP/co-integrator protein (P/CIP) and the activator of retinoid receptors (ACTR)"

/codon_start=1

/product="receptor-associated coactivator 3"

/protein_id="RAC31663.1"

/db_xref="PID:92318006"

/db_xref="GI:2318006"

/translation="MSGIGENLDPASDRKRRLPCDTPQGLTSGEKRRRQESRY
IEELAEIDLSNLDNFNVKDKCAILKETROIROIKQKGTISNDDVQADYSS
TGQVIDKSLGPDLLQALDGLFVNVREANIVFVSENVTOYLOKQEDLVNTSVNI
LHEEDKDFLKNLPSKTVNGSVNTNETQKSHTEFCNMLKTPHDLIEDINASPENR
QRYTEMCFALSPRAMEGEDIQSCMICVARRITTGERTFFSPNPFITRHDLSEK
VNIIDTNSLRSGMPGEDIIRICIQRFSLNDQSWSKRHYQAYLNGHATPVYR
FSLADGTIVTAQTSKLFNRPVNDHRGFVSTHFLQREONGYRPNPNPVGQIRPMA
GCNSVGMGMSNQGLOPSSRAYGLADPSTTGOMSGARYGSSNTASLTPOGPMOS
PSSYONNGLNMSPPHSGPLAPNOQNMIMSPRNRGSPKIASHOFSPVAGVHSPWA
SSGNTGNHSSSSLSALQISEGVGTSLTSLSPGPKLDNPNMNIQPSKVNQD
SKSPGTYCQNPVSMSCQNSRDLSDKSESSVEGAENQRPLESKGRKLLQL
LTCSSDRHSSSLNPLDSCKESSVSITSPGVSSTSGVSSVSNMHSLLQEKH
RILHLKNGNSPAEVAKITAETAKDTSITSCGDGVYKQQLSPKKNENALLRY
LLDRDPDSALSKLOPQVEGVNKMSCQTSSTIPSSQEKDPKIKPTSEEGSDLD
NLDAIILGDTSSDFNLSISNGSHLCTKQVFOGTNSLGLKSSQSVQSIIRPVNRAV
SLDSPSVSGSPVKNLSAPPLPKPQMLGNPRMDSQENYSSMGSPNRYTVQT
PSSDWDGLPSKAGRMPSNMSGNRPFGDYNTSLPALGGSITPLPLNSNIPGAR
PVLOQOQMLQRPGEIIPMGMAVYQGAASNOLGSPDMLSMGVSQHTGNRPL
RNSLDDLVGPPNSGDSERALLDLHTLSNTDGLTLEIDRALGIPELVNGQAL
EPKQDAFGQEAAYVMQKAGLYGQYPAQPGMPGFFHLQGGSPSPSNMNMNOGQ
NFPQGMHPRANIRPTNPKOLRMOLQORLOGOQFLNOSROALEUKMNPATAGGA
VMRPMQOQGFLENAQVAORSRELLSHFERQVAVNMQOQOQOQOQOQOQOQOQ
OQOQOQOQOQAFSPPNPTAFSDMLLAGTMTFQAPQPFPIQPNMGQOQDPAFGR
VSPPNAMMSRMGFSQNPQMHPQAAAIYQSEMKGWPSGNLARNSSFSQOQFAHOG
NPAYYSVHMNGSSGHGMQNMNPMPSGMPMGPDQKYC"

1354 a 1048 -c 1043 g 1050 t

ORIGIN

Query Match 99.48; Score 4468.2; DB 11; Length 4495;
Best Local Similarity 99.88; Pred. No. 0;
Matches 4486; Conservative 0; Mismatches 3; Indels 6; Gaps 1;
QY 1 GCTGGATGCTGATGATATTAACAGTGGATTAGAGAAACCTTGGATCCACTGGC 60
DB 1 GCTGGATGCTGATGATATTAACAGTGGATTAGAGAAACCTTGGATCCACTGGC 60
QY 61 AGCCAGTTGCTGATGATATTAACAGTGGATTAGAGAAACCTTGGATCCACTGGC 120
DB 61 AGCCAGTTGCTGATGATATTAACAGTGGATTAGAGAAACCTTGGATCCACTGGC 120
QY 121 CAGTGAATTCAGAAACGCAATATGCCATGTGATCTCCAGAGCAAGGCTTACCTGCG 180
DB 121 CAGTGAATTCAGAAACGCAATATGCCATGTGATCTCCAGAGCAAGGCTTACCTGCG 180
QY 181 TGGTGAACACGGACGGACGAGGAAGTAAATATATTGAAGATGGCTGACCTGAT 240
DB 181 TGGTGAACACGGACGGACGAGGAAGTAAATATATTGAAGATGGCTGACCTGAT 240
QY 241 ATCTGCCAATCTTAGTGTATTTGACAATTTCAATGTCAACACGAGATAAATGCGGATTT 300

Db 1371 GAGCAGAGGGCCCTATGGCTTGGCAGACCCCTAGCACACAGGGCAGATGAGTGGAGCTAG 1380
Qy 1381 GTATGGGGTTCCAGTAACATAGCTTCAATGACCCCTGGGCCAGGCATGCAATCACCATC 1440
Db 1381 GTATGGGGTTCCAGTAACATAGCTTCAATGACCCCTGGGCCAGGCATGCAATCACCATC 1440
Qy 1441 TTCTTACCAGAACAACTATGGCTCAACATGAGTAGCCGCCCCACCATGGAGTCTCTGG 1500
Db 1441 TTCTTACCAGAACAACTATGGCTCAACATGAGTAGCCGCCCCACCATGGAGTCTCTGG 1500
Qy 1501 TCTTGGCCCAACACAGAGATATCATGATTTCTCTGTAATCGTGGAGTCCAAAGAT 1560
Db 1501 TCTTGGCCCAACACAGAGATATCATGATTTCTCTGTAATCGTGGAGTCCAAAGAT 1560
Qy 1561 AGCCTCAATCAGATTTCTCTGTTGACAGTGTGCACCTCTCCCATGGCATCTTCTGGCAA 1620
Db 1561 AGCCTCAATCAGATTTCTCTGTTGACAGTGTGCACCTCTCCCATGGCATCTTCTGGCAA 1620
Qy 1621 TACTGGGNACACAGCTTTTCCAGAGCTCTCTCAGTGGCCCTGCAAGCCATCAGTGAAG 1680
Db 1621 TACTGGGNACACAGCTTTTCCAGAGCTCTCTCAGTGGCCCTGCAAGCCATCAGTGAAG 1680
Qy 1681 TGTGGGACTTCCCTTTATCTACTCTCATCACAGGCCCCAAATTTGGATACTCTCC 1740
Db 1681 TGTGGGACTTCCCTTTATCTACTCTCATCACAGGCCCCAAATTTGGATACTCTCC 1740
Qy 1741 CAATATGAATATTACCCCAACCAAGTAAGTAAGCAATCAGGATTCCTCTGGG 1800
Db 1741 CAATATGAATATTACCCCAACCAAGTAAGTAAGCAATCAGGATTCCTCTGGG 1800
Qy 1801 CTTTATTTGACCAAAATCCAGTGGAGATTCATGTGTGTCAGTCAATATACAGAGATCA 1860
Db 1801 CTTTATTTGACCAAAATCCAGTGGAGTTCATGTGTGTCAGTCAATATACAGAGATCA 1860
Qy 1861 CCTCAGTGACAAAGAAAGTAAGGAGAGCAGTGTGGAGGGCAGAGATCAAGGGGTCC 1920
Db 1861 CCTCAGTGACAAAGAAAGTAAGGAGAGCAGTGTGGAGGGCAGAGATCAAGGGGTCC 1920
Qy 1921 TTTGAAAGCAAGGTCATAAAAAATTACTGCAGTTACTTACCTGTCTTCTGATGACCG 1980
Db 1921 TTTGAAAGCAAGGTCATAAAAAATTACTGCAGTTACTTACCTGTCTTCTGATGACCG 1980
Qy 1981 GGGTCATTCCTCTGACCACTCCCTCAGATTCAGATTCAGATTCCTCTGTTAG 2040
Db 1981 GGGTCATTCCTCTGACCACTCCCTCAGATTCAGATTCAGATTCCTCTGTTAG 2040
Qy 2041 TGTCAACAGCCCTCTGGAGTCTCTCTCTACATCTGGAGGATCTCTTACATCAA 2100
Db 2041 TGTCAACAGCCCTCTGGAGTCTCTCTCTACATCTGGAGGATCTCTTACATCAA 2100
Qy 2101 TATGCATGGGTCACTGTTTACAGAGAACACCGGATTTTGACAGATTTGTCGAGAATGG 2160
Db 2101 TATGCATGGGTCACTGTTTACAGAGAACACCGGATTTTGACAGATTTGTCGAGAATGG 2160
Qy 2161 GAATTCACCACTGAGGTAGCCAGATTTACTGACAGCCACTGGGAAGACACAGCAG 2220
Db 2161 GAATTCACCACTGAGGTAGCCAGATTTACTGACAGCCACTGGGAAGACACAGCAG 2220
Qy 2221 TATACCTCTTGTGGGACGGAATGTTGTCAAGCAGAGCAGCTAAGTCTTAAGAGAA 2280
Db 2221 TATACCTCTTGTGGGACGGAATGTTGTCAAGCAGAGCAGCTAAGTCTTAAGAGAA 2280
Qy 2281 GGAGAAATGCACTTCTTAGATACCTCTGACAGGGATGATCTAGTATGACCTCTC 2340
Db 2281 GGAGAAATGCACTTCTTAGATACCTCTGACAGGGATGATCTAGTATGACCTCTC 2340
Qy 2341 TAAAGAACTACAGCCCAAGTGGAGGAGTGGACAATAAATGATCAGTGCACCACTC 2400
Db 2341 TAAAGAACTACAGCCCAAGTGGAGGAGTGGACAATAAATGATCAGTGCACCACTC 2400
Qy 2401 CACCATCTAGTCAAGTCAAGGAAAGACCCCTAAATTAAGCAGAGACAAAGTGAAGA 2460
Db 2401 CACCATCTAGTCAAGTCAAGGAAAGACCCCTAAATTAAGCAGAGACAAAGTGAAGA 2460

Qy 2461 GGGATCTGGAGACTTGGATAATCTAGATGCTATTCCTGGTATCTGACTAGTCTGACTT 2520
Db 2461 GGGATCTGGAGACTTGGATAATCTAGATGCTATTCCTGGTATCTGACTAGTCTGACTT 2520
Qy 2521 TTACAATAATTCATATCCTCAATGCTAGTATCTCTGGGACTAAGCAACAGGTGTTTCA 2580
Db 2521 TTACAATAATTCATATCCTCAATGCTAGTATCTCTGGGACTAAGCAACAGGTGTTTCA 2580
Qy 2581 AGGAACATAATTCCTGGTGGTAAAGTTTCAAGTCTGACAGTCTATTCCTCTCCATA 2640
Db 2581 AGGAACATAATTCCTGGTGGTAAAGTTTCAAGTCTGACAGTCTATTCCTCTCCATA 2640
Qy 2641 TAACCGAGCAGTGTCTCTGGATAGCCCTGTTCTGTTGGCTCAAGTCTCCAGTAAAAA 2700
Db 2641 TAACCGAGCAGTGTCTCTGGATAGCCCTGTTCTGTTGGCTCAAGTCTCCAGTAAAAA 2700
Qy 2701 TATCAGTGTCTTCCCATGTTTACCAAGCAACCCATGTTGGGTGGGAATCCAAAGATGAT 2760
Db 2701 TATCAGTGTCTTCCCATGTTTACCAAGCAACCCATGTTGGGTGGGAATCCAAAGATGAT 2760
Qy 2761 GGATAGTCAGGAAATTTATGGCTCAAGTATGGGTGGGCAAAACCGAATGTGACTGTGAC 2820
Db 2761 GGATAGTCAGGAAATTTATGGCTCAAGTATGGGTGGGCAAAACCGAATGTGACTGTGAC 2820
Qy 2821 TCAGACTCTTCTCTCAGGAGACTGGGCTTACCAAACTCAAAAGCCGCGCAGAAATGGAACC 2880
Db 2821 TCAGACTCTTCTCTCAGGAGACTGGGCTTACCAAACTCAAAAGCCGCGCAGAAATGGAACC 2880
Qy 2881 TATGAATTCAACTCCATGGGAAGACCCAGGAGAGATTAATTAATCTTCTTACCCAGACC 2940
Db 2881 TATGAATTCAACTCCATGGGAAGACCCAGGAGAGATTAATTAATCTTCTTACCCAGACC 2940
Qy 2941 TGCAGTGGGTGGCTCTATTCACATTTCCCTCTCTGCTTAATAGCATACCAAGTGGAG 3000
Db 2941 TGCAGTGGGTGGCTCTATTCACATTTCCCTCTCTGCTTAATAGCATACCAAGTGGAG 3000
Qy 3001 ACCAGTATTCACAGCAGCAGCAGATGCTTCAAAATGAGGCTTGGTAAATCCCATGGG 3060
Db 3001 ACCAGTATTCACAGCAGCAGCAGATGCTTCAAAATGAGGCTTGGTAAATCCCATGGG 3060
Qy 3061 AATGGGGCTTAATCCCTATGGCCAAAGCAGCAGATCTAACCACCTGGTCTTGGCCGGA 3120
Db 3061 AATGGGGCTTAATCCCTATGGCCAAAGCAGCAGATCTAACCACCTGGTCTTGGCCGGA 3120
Qy 3121 TGGCATGTTGTCATGGAACAAGTTCTCTGACACTCAAAATAGGCTCTTCTTAGAA 3180
Db 3121 TGGCATGTTGTCATGGAACAAGTTCTCTGACACTCAAAATAGGCTCTTCTTAGAA 3180
Qy 3181 TTCCCTGGATGATCTTGTGGGCCACCTTCCAACTGGAAGCCAGAGTACGAAAGAGC 3240
Db 3181 TTCCCTGGATGATCTTGTGGGCCACCTTCCAACTGGAAGCCAGAGTACGAAAGAGC 3240
Qy 3241 ATTATTTGGACAGCTGCACACTCTTCTCAGCAACACAGATGCGACAGCCCTGGAAGAA 3300
Db 3241 ATTATTTGGACAGCTGCACACTCTTCTCAGCAACACAGATGCGACAGCCCTGGAAGAA 3300
Qy 3301 TGACAGAGCTTGGGCATCTCTGAACTTGTCAATCAGGACAGGCATTAGAGCCCAACA 3360
Db 3301 TGACAGAGCTTGGGCATCTCTGAACTTGTCAATCAGGACAGGCATTAGAGCCCAACA 3360
Qy 3361 GGATGCTTTTCAAGGCCAAGAACGACAGATTAATGATGATCAGAGGCGAGGATTAATGG 3420
Db 3361 GGATGCTTTTCAAGGCCAAGAACGACAGATTAATGATGATCAGAGGCGAGGATTAATGG 3420
Qy 3421 ACAGACATACCCAGCAGAGGGCCCTCCAATGCAAGAGGCTTTCATCTTCAGGACAAATC 3480
Db 3421 ACAGACATACCCAGCAGAGGGCCCTCCAATGCAAGAGGCTTTCATCTTCAGGACAAATC 3480
Qy 3481 ACCATCTTTTAACTCTATGATGAATCAGATGAACCCAGCAAGCAATTTCTCTCCTCAAG 3540
Db 3481 ACCATCTTTTAACTCTATGATGAATCAGATGAACCCAGCAAGCAATTTCTCTCCTCAAG 3540

QY 121 CAGTGATTCAGAAAACGCAAAATGGCCATGTGATCTCCAGGACGAAGGCTTTACTCTGCAG 180
DB 236 CAGTGATTCAGAAAACGCAAAATGGCCATGTGATCTCCAGGACGAAGGCTTTACTCTGCAG 295
QY 181 TGGTGAAGAACGGAGACGGGAGCAGGAAGTAATATATTCAAGAATTTGGCTGAGCTGAT 240
DB 296 TGGTGAAGAACGGAGACGGGAGCAGGAAGTAATATATTCAAGAATTTGGCTGAGCTGAT 355
QY 241 ATCTGCCAATCTTAGTGATATTGCAATTTTCAATTTGTCAAACAGATTAATTTGGGATTTT 300
DB 356 ATCTGCCAATCTTAGTGATATTGCAATTTTCAATTTGTCAAACAGATTAATTTGGGATTTT 415
QY 301 AAAGGAACAGTAGACAGATACGTCAATATAAGAGAGCAAGGAAGAACTATTTCATGA 360
DB 416 AAAGGAACAGTAGACAGATACGTCAATATAAGAGAGCAAGGAAGAACTATTTCATGA 475
QY 361 TGATGATGTTCAAAAAGCCGATGTATCTCTACAGGCGAGGAGTTATTGATAAGACTC 420
DB 476 TGATGATGTTCAAAAAGCCGATGTATCTCTACAGGCGAGGAGTTATTGATAAGACTC 535
QY 421 CTTAGGACCGCTTTTACTTCAGGCATTTGGATGGTTTCTTATTTGGTGAATCGAGAGGC 480
DB 536 CTTAGGACCGCTTTTACTTCAGGCATTTGGATGGTTTCTTATTTGGTGAATCGAGAGGC 595
QY 481 AAACATTGATTTGTATCAGAAAATGTCACACAATACCTGCAATATAAGCAAGAGACCT 540
DB 596 AAACATTGATTTGTATCAGAAAATGTCACACAATACCTGCAATATAAGCAAGAGACCT 655
QY 541 GGTAAACACAGTGTTCACATATCTTACATGAAGAGACAGAAAGGATTTTCTTAAGAA 600
DB 656 GGTAAACACAGTGTTCACATATCTTACATGAAGAGACAGAAAGGATTTTCTTAAGAA 715
QY 601 TTTACCAAAATCTACAGTTAATGGAGTTTCTCTGGCAAAATGAGCCCAACACAAAAAG 660
DB 716 TTTACCAAAATCTACAGTTAATGGAGTTTCTCTGGCAAAATGAGCCCAACACAAAAAG 775
QY 661 CCATACATTTAATTTGCCGTATGTTGATGAAGAACACACATGATATCTCGAAGACATAAA 720
DB 776 CCATACATTTAATTTGCCGTATGTTGATGAAGAACACACATGATATCTCGAAGACATAAA 835
QY 721 CGCCAGTCTGAATGCGCCAGAGATATGAACAATGCAAGTGTCTGACCCCTGTCAGCC 780
DB 836 CGCCAGTCTGAATGCGCCAGAGATATGAACAATGCAAGTGTCTGACCCCTGTCAGCC 895
QY 781 ACAGACTATGATGAGGAAGGGAAGATTTGCAATCTGTATGATCTGTGTGGCACGCGC 840
DB 896 ACAGACTATGATGAGGAAGGGAAGATTTGCAATCTGTATGATCTGTGTGGCACGCGC 955
QY 841 CATTACTACAGGAGAAAGACATTTCCATCAAAACCCCTGAGAGCTTTATTACCAGACATGA 900
DB 956 CATTACTACAGGAGAAAGACATTTCCATCAAAACCCCTGAGAGCTTTATTACCAGACATGA 1015
QY 901 TCTTTCAGGAAGGTTGTCAATATAGATACAAATTCAGTACATCTCCATGAGCCCTGG 960
DB 1016 TCTTTCAGGAAGGTTGTCAATATAGATACAAATTCAGTACATCTCCATGAGCCCTGG 1075
QY 961 CTTTGAAGATATAATCCGAAGGTTATTTCAGAGATTTTATTAGTCTTAATGATGSCAGTC 1020
DB 1076 CTTTGAAGATATAATCCGAAGGTTATTTCAGAGATTTTATTAGTCTTAATGATGSCAGTC 1135
QY 1021 ATGTTCCAGAAACGTCATCTAAGAGCTTATCTTAATGGCCATGAGAAACCCAGT 1080
DB 1136 ATGTTCCAGAAACGTCATCTAAGAGCTTATCTTAATGGCCATGAGAAACCCAGT 1195
QY 1081 ATATCGATTCTCGTTGGCTGATGGAACATATAGTACCTGACAGACAAAGCAAACTCTT 1140
DB 1196 ATATCGATTCTCGTTGGCTGATGGAACATATAGTACCTGACAGACAAAGCAAACTCTT 1255
QY 1141 CCGAAATCCTGTAACAATGATCGACATGGCTTTGCTCAACCCACTTCTTCAGAGAGA 1200
DB 1256 CCGAAATCCTGTAACAATGATCGACATGGCTTTGCTCAACCCACTTCTTCAGAGAGA 1315

QY 1201 ACAGAAATGGATATATAGACCAAAACCCAAATCTCTGTTGGACAGGGATTTAGACCACCTATGGC 1260
DB 1316 ACAGAAATGGATATATAGACCAAAACCCAAATCTCTGTTGGACAGGGATTTAGACCACCTATGGC 1375
QY 1361 TGGATGCAACAGTTTCGGTAGCGGCATGAGTATCTGCGCAAAACCAAGCTTTACAGATGCC 1320
DB 1376 TGGATGCAACAGTTTCGGTAGCGGCATGAGTATCTGCGCAAAACCAAGCTTTACAGATGCC 1435
QY 1321 GAGCAGAGGSCCTTATGGCTTGGCAGACCCCTAGCACACAGGGCAGATGAGTGGAGCTAG 1380
DB 1436 GAGCAGAGGSCCTTATGGCTTGGCAGACCCCTAGCACACAGGGCAGATGAGTGGAGCTAG 1495
QY 1381 GTATGGGGTTCAGTAAACATAGCTTCATTTGACCCCTGGGCGAGCATGCAATCACCATC 1440
DB 1496 GTATGGGGTTCAGTAAACATAGCTTCATTTGACCCCTGGGCGAGCATGCAATCACCATC 1555
QY 1441 TTCTTACAGAACCAACATATGGCTCAACATGAGTAGCCCCCACATGGAGTCTCTGG 1500
DB 1556 TTCTTACAGAACCAACATATGGCTCAACATGAGTAGCCCCCACATGGAGTCTCTGG 1615
QY 1501 TCTTGGCCCAACAGCAGATATCATGATTTCTCTCGTAATCGTGGAGTCCAAAGAT 1560
DB 1616 TCTTGGCCCAACAGCAGATATCATGATTTCTCTCGTAATCGTGGAGTCCAAAGAT 1675
QY 1561 AGCCTCACATCAGTTTCTCTGTTGCAAGTGTGCATCTCTCCATGCGCATCTTCTGGCAA 1620
DB 1676 AGCCTCACATCAGTTTCTCTGTTGCAAGTGTGCATCTCTCCATGCGCATCTTCTGGCAA 1735
QY 1621 TACTGGGAACCAACAGCTTTTCCAGCAGCTCTCTCAGTGGCCCTGCAAGCCATCAGTGAAG 1680
DB 1736 TACTGGGAACCAACAGCTTTTCCAGCAGCTCTCTCAGTGGCCCTGCAAGCCATCAGTGAAG 1795
QY 1681 TGTGGGACTTCCCTTTTATCTACTCTGTCTATCACCAGGCCCAATTTGGATTAACCTCC 1740
DB 1796 TGTGGGACTTCCCTTTTATCTACTCTGTCTATCACCAGGCCCAATTTGGATTAACCTCC 1855
QY 1741 CAATATGAATATATACCAACCAAGTAAGTAAGCAATCAGGATTTCCAAGAGTCTCTGG 1800
DB 1856 CAATATGAATATATACCAACCAAGTAAGTAAGCAATCAGGATTTCCAAGAGTCTCTGG 1915
QY 1801 CTTTATTTGGCCACCAAAATCCAGTGGAGAGTTCAATGTGTGTCAGTCAATATCAGAGATCA 1860
DB 1916 CTTTATTTGGCCACCAAAATCCAGTGGAGAGTTCAATGTGTGTCAGTCAATATCAGAGATCA 1975
QY 1861 CCTCAGTGACAAAAGTAAGAGAGCAGTGTGTGAGGGGCGAGAGATCAAAAGGGTCC 1920
DB 1976 CCTCAGTGACAAAAGTAAGAGAGCAGTGTGTGAGGGGCGAGAGATCAAAAGGGTCC 2035
QY 1921 TTTGGAAGCAAAAGGTCATAAAAAATTTACTGCACTTACTTTACCTGTTCTTCTGATGACC 1980
DB 2036 TTTGGAAGCAAAAGGTCATAAAAAATTTACTGCACTTACTTTACCTGTTCTTCTGATGACC 2095
QY 1981 GGGTCATCTCTCTTGACCAACTCCCCCTAGATTCAGTTGTAAGAATCTTCTGTTAG 2040
DB 2096 GGGTCATCTCTCTTGACCAACTCCCCCTAGATTCAGTTGTAAGAATCTTCTGTTAG 2155
QY 2041 TGTCAACAGCCCTCTGGAGTCTCTCTCTACATCTGGAGAGTATCTCTACATCAA 2100
DB 2156 TGTCAACAGCCCTCTGGAGTCTCTCTCTACATCTGGAGAGTATCTCTACATCAA 2215
QY 2101 TATGATGGGTCACTGTTTACAAGAGAACCCGGATTTTGCAAGTGTGTCAGAGATGG 2160
DB 2216 TATGATGGGTCACTGTTTACAAGAGAACCCGGATTTTGCAAGTGTGTCAGAGATGG 2275
QY 2161 GAATTCACAGCTGAGGTAGCCAAAGTATTCTGCAACAGCCACTGGGAAGACACACGAG 2220
DB 2276 GAATTCACAGCTGAGGTAGCCAAAGTATTCTGCAAGAGCCACTGGGAAGACACACGAG 2335
QY 2221 TATAACTTCTTGTGGGACCGAAATGTTGCAAGCAGGAGCAGCTTAAGTCTCTAAGAGAA 2280
DB 2336 TATAACTTCTTGTGGGACCGAAATGTTGCAAGCAGGAGCAGCTTAAGTCTCTAAGAGAA 2395
QY 2281 GGAGATAATATGCACTTCTTAGATACCTGCTGGACAGGGATGATCCTAGTATGATGCACTCTC 2340

|||||
Db 2396 GGAGAAATATGCATTCTTATAGATACCTGCTGGACAGGATGATCCTAGTGATGCACTCTC 2455
|||||
Qy 2341 TAAAGAACTACACCCCAAGTGAAGAGTGGACATAATAATAGTCAAGTGCACAGCTC 2400
|||||
Db 2456 TAAAGAACTACACCCCAAGTGAAGAGTGGATATAATAATAGTCAAGTGCACAGCTC 2515
|||||
Qy 2401 CACCATCTCTAGTCAAGTCAAGTCAAGAAAGACCTTAAATTAAGACAGAGACAGTGAAGA 2460
|||||
Db 2516 CACCATCTCTAGTCAAGTCAAGAAAGACCTTAAATTAAGACAGAGACAGTGAAGA 2575
|||||
Qy 2461 GGGATCTGGAGACTTGGATTAATCTAGATGCTATTCTTGGTGATCTGACTAGTTCTGACTT 2520
|||||
Db 2576 GGGATCTGGAGACTTGGATTAATCTAGATGCTATTCTTGGTGATCTGACTAGTTCTGACTT 2635
|||||
Qy 2521 TTACAATAATTCATATCTCAATGGTAGTCAATCTGAGCTCTATTGCTCTCCATA 2580
|||||
Db 2636 TTACAATAATTCATATCTCAATGGTAGTCAATCTGAGGACTTAAGCAACAGGTGTTCA 2695
|||||
Qy 2581 AGGAACCTAATCTCTGGGTTTGAAGAGTTTCAACAGTCTGTGACGCTATTGCTCTCCATA 2640
|||||
Db 2696 AGGAACCTAATCTCTGGGTTTGAAGAGTTTCAACAGTCTGTGACGCTATTGCTCTCCATA 2755
|||||
Qy 2641 TAACCGAGCAGTCTCTGATAGCCCTGTTTCTGTTGGCTCAAGCTCTCAGTAAATAA 2700
|||||
Db 2756 TAACCGAGCAGTCTCTGATAGCCCTGTTTCTGTTGGCTCAAGCTCTCAGTAAATAA 2815
|||||
Qy 2701 TATCAGTCTTCCCATGTTTACAAAGCAACCCATGTTGGGTGGGAATCCAAAGATGAT 2760
|||||
Db 2816 TATCAGTCTTCCCATGTTTACAAAGCAACCCATGTTGGGTGGGAATCCAAAGATGAT 2875
|||||
Qy 2761 GGATAGTCAGAAATATATGGCTCAAGTATGGTGGGCCAAACCGAAATGTGACTGTGAC 2820
|||||
Db 2876 GGATAGTCAGAAATATATGGCTCAAGTATGGTGGGCCAAACCGAAATGTGACTGTGAC 2935
|||||
Qy 2821 TCAGACTCTCTCCAGGAGACTGGGCTTACCAAACTCAAGGCCGGCAGAGATGGAACC 2880
|||||
Db 2936 TCAGACTCTCTCCAGGAGACTGGGCTTACCAAACTCAAGGCCGGCAGAGATGGAACC 2995
|||||
Qy 2881 TATGAATTCAACTCCATGGGAAGACAGGAGAGATTAATACTCTTTACCCAGACC 2940
|||||
Db 2996 TATGAATTCAACTCCATGGGAAGACAGGAGAGATTAATACTCTTTACCCAGACC 3055
|||||
Qy 2941 TGCACTGGGTGGCTCTATTCCTCCATGCTCTTCGCTCTTAATAGATACCAAGTGGAG 3000
|||||
Db 3056 TGCACTGGGTGGCTCTATTCCTCCATGCTCTTCGCTCTTAATAGATACCAAGTGGAG 3115
|||||
Qy 3001 ACCAGTATTCGAACAGCAGCAGATGCTTCAATAGAGCCCTGGTGAATCCCCATGGG 3060
|||||
Db 3116 ACCAGTATTCGAACAGCAGCAGATGCTTCAATAGAGCCCTGGTGAATCCCCATGGG 3175
|||||
Qy 3061 ANTGGGGCTTAATCCCTATGGCCAAAGCAGCAGATCTAACCACTGGGTCTGGGCCGA 3120
|||||
Db 3176 ANTGGGGCTTAATCCCTATGGCCAAAGCAGCAGATCTAACCACTGGGTCTGGGCCGA 3235
|||||
Qy 3121 TGGCATCTTGTCCATGAACAAAGTTCTCATGCACTCAAAATAGGCTCTCTCTTAGGAA 3180
|||||
Db 3236 TGGCATCTTGTCCATGAACAAAGTTCTCATGCACTCAAAATAGGCTCTCTCTTAGGAA 3295
|||||
Qy 3181 TTCCCTGGATGATCTTGTGGGCCACCTTCCAACTGGGAAGGCCAGAGTGAACAAAGAGC 3240
|||||
Db 3296 TTCCCTGGATGATCTTGTGGGCCACCTTCCAACTGGGAAGGCCAGAGTGAACAAAGAGC 3355
|||||
Qy 3241 ATTATTGGACAGTGGCAGACTCTTCTCAGCAACACAGATGCGACAGCCCTGGAAGAAAT 3300
|||||
Db 3356 ATTATTGGACAGTGGCAGACTCTTCTCAGCAACACAGATGCGACAGCCCTGGAAGAAAT 3415
|||||
Qy 3301 TGACAGAGCTTGGGCAATCTTGAACCTTGTCAATCAGGAGCAGCATTAGAGCCCAACA 3360
|||||
Db 3416 TGACAGAGCTTGGGCAATCTTGAACCTTGTCAATCAGGAGCAGCATTAGAGCCCAACA 3475
|||||
Qy 3361 GGATGCTTCCAAAGGCCAAGACAGCAGTAAATGATGGATCAGAAGSCAGGATTATGG 3420
|||||

Db 3476 GGATGCTTTCCAAAGGCCAAAGACAGCAGTAAATGATGGATCAGAAGCAGGATTATATGG 3535
Qy 3421 ACAGACATACCCAGCAGCAGGGCCCTCCAATGCAAGAGAGCTTTTCATCTTCAGGGAATC 3480
|||||
Db 3536 ACAGACATACCCAGCAGCAGGGCCCTCCAATGCAAGAGAGCTTTTCATCTTCAGGGAATC 3595
|||||
Qy 3481 ACCATCTTTAACTCTATGATGAATCAGATGAACAGCAGAGGCAATTTTCCTCTCCAAG 3540
|||||
Db 3596 ACCATCTTTAACTCTATGATGAATCAGATGAACAGCAGAGGCAATTTTCCTCTCCAAG 3655
|||||
Qy 3541 ATGCAACCCAGCAGCAGCAATCATGAGACCCCGGACAAACACCCCAAGCAACTTTAGAAT 3600
|||||
Db 3656 AATGCAACCCAGCAGCAGCAATCATGAGACCCCGGACAAACACCCCAAGCAACTTTAGAAT 3715
|||||
Qy 3601 GCAGCTTTCAGCAGAGGCTGAGGGCCAGCAGTTTGAATCAGAGCCGACAGGCACTTGA 3660
|||||
Db 3716 GCAGCTTTCAGCAGAGGCTGAGGGCCAGCAGTTTGAATCAGAGCCGACAGGCACTTGA 3775
|||||
Qy 3661 ATTGAAATGGAACCCCTACTGCTGGTGTGCTGGGTGATGAGGCTTATGATGAGCC 3720
|||||
Db 3776 ATTGAAATGGAACCCCTACTGCTGGTGTGCTGGGTGATGAGGCTTATGATGAGCC 3835
|||||
Qy 3721 CCAGCAGGCTTTTCTTAATGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3780
|||||
Db 3836 CCAGCAGGCTTTTCTTAATGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3895
|||||
Qy 3781 TCACCTCCACACACAGAGGCTGCTATGATGATGAGCAG-----CAGCA 3825
|||||
Db 3896 TCACCTCCACACACAGAGGCTGCTATGATGATGAGCAGCAGCAGCAGCAACAGCA 3955
|||||
Qy 3826 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3885
|||||
Db 3956 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4015
|||||
Qy 3886 AACCCAGGCTTTCAGCCACCTCCTTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3945
|||||
Db 4016 AACCCAGGCTTTCAGCCACCTCCTTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4075
|||||
Qy 3946 GGCAGACCCCAATGCCACAGCTCCTCCGCAACAGTTTCCATATCAACCAATATG 4005
|||||
Db 4076 GGCAGACCCCAATGCCACAGCTCCTCCGCAACAGTTTCCATATCAACCAATATG 4135
|||||
Qy 4006 AATGGCAACCAACACAGATCCAGCTTTCGTCAGTGTCTAGTCTCCCAATCAATGAT 4065
|||||
Db 4136 AATGGCAACCAACACAGATCCAGCTTTCGTCAGTGTCTAGTCTCCCAATCAATGAT 4195
|||||
Qy 4066 GTGCTCAAGAAATGGGTCTCCAGAAATCCCATGATGCAACACCCCGCAGCTGCATCCAT 4125
|||||
Db 4196 GTGCTCAAGAAATGGGTCTCCAGAAATCCCATGATGCAACACCCCGCAGCTGCATCCAT 4255
|||||
Qy 4126 CTATCAGTCTCAGAAATGAAGGCTGGCCATCAGGAAATTTGGCCAGGACAGCTCCTT 4185
|||||
Db 4256 CTATCAGTCTCAGAAATGAAGGCTGGCCATCAGGAAATTTGGCCAGGACAGCTCCTT 4315
|||||
Qy 4186 TTCCCAAGCAGCAGTTTGGCCACCCAGGGGAATCCTGCAAGTGTATAGTATGGTGACATGAA 4245
|||||
Db 4316 TTCCCAAGCAGCAGTTTGGCCACCCAGGGGAATCCTGCAAGTGTATAGTATGGTGACATGAA 4375
|||||
Qy 4246 TGGCAGCAGTGTACATGGGACAGATGAACATGAACCCCATGCCCATGTCTGGCATGCC 4305
|||||
Db 4376 TGGCAGCAGTGTGTACATGGGACAGATGAACATGAACCCCATGCCCATGTCTGGCATGCC 4435
|||||
Qy 4306 TATGGGTCTCTGATCAGAAATACTGCTGACATCTCTGCACAGGACCTCTTAAGGAACCA 4365
|||||
Db 4436 TATGGGTCTCTGATCAGAAATACTGCTGACATCTCTGCACAGGACCTCTTAAGGAACCA 4495
|||||
Qy 4366 CTGTACAAATGACACTGCACTAGGATTTATGGGAAGGAATCATTTGTCAGGACATCCATC 4425
|||||
Db 4496 CTGTACAAATGACACTGCACTAGGATTTATGGGAAGGAATCATTTGTCAGGACATCCATC 4555
|||||
Qy 4426 TTGGAAAGAAAGGACAGCTTTGAGCTCCCATCAAGGGTATTTTAAAGTGTATGTTGAGC 4485
|||||
Db 4556 TTGGAAAGAAAGGACAGCTTTGAGCTCCCATCAAGGGTATTTTAAAGTGTATGTTGAGC 4615
|||||

[illegible]

QY 1021 ATGGTCCCAAGAACGTCACATCAAGAGCTTATCTTAATGGCCATGAGAAACCCAGT 1080
DB 1127 ATGGTCCCAAGAACGTCACATCAAGAGCTTATCTTAATGGCCATGAGAAACCCAGT 1186
QY 1081 ATATCGATTCTCGTTGGCTGATGGAACCTATAGTACTGCACAGACAAAAGCAAACTCTT 1140
DB 1187 ATATCGATTCTCGTTGGCTGATGGAACCTATAGTACTGCACAGACAAAAGCAAACTCTT 1246
QY 1141 CGAAATCTGTAAACAAATGATCGACATGGCTTTGTCTCAACCCACTTCCCTTCAGAGAGA 1200
DB 1247 CGAAATCTGTAAACAAATGATCGACATGGCTTTGTCTCAACCCACTTCCCTTCAGAGAGA 1306
QY 1201 ACAGAAATGATATAGACAAACCCAAATCTCTTGGCAAGGATATAGACCACTTATGGC 1260
DB 1307 ACAGAAATGATATAGACAAACCCAAATCTCTTGGCAAGGATATAGACCACTTATGGC 1366
QY 1261 TGGATGCAACAGTTCCGTAGGGCGCATGAGTATGTGCGCAAAACCAAGGCTTACAGATGCC 1320
DB 1367 TGGATGCAACAGTTCCGTAGGGCGCATGAGTATGTGCGCAAAACCAAGGCTTACAGATGCC 1426
QY 1321 GAGCAGAGGGCTATGGCTTGGCAGACCTAGCACCACAGGCGAGATGAGTGGAGCTAG 1380
DB 1427 GAGCAGAGGGCTATGGCTTGGCAGACCTAGCACCACAGGCGAGATGAGTGGAGCTAG 1486
QY 1381 GTATGGGGTTCCAGTAACATAGCTTTCATTGACCCCTGGCGCAGGCATGCAATCACCATC 1440
DB 1487 GTATGGGGTTCCAGTAACATAGCTTTCATTGACCCCTGGCGCAGGCATGCAATCACCATC 1546
QY 1441 TTCTTACCAGAACAACTATGGGCTCAACATGAGTAGCCGCCCCACATGGGAGTCTCTGG 1500
DB 1547 TTCTTACCAGAACAACTATGGGCTCAACATGAGTAGCCGCCCCACATGGGAGTCTCTGG 1606
QY 1501 TCTTCCGCCCAACACAGCAATATCATGTTCTCTCGTATCTGGAGTCCCAAGAT 1560
DB 1607 TCTTCCGCCCAACACAGCAATATCATGTTCTCTCGTATCTGGAGTCCCAAGAT 1666
QY 1561 AGCCTCACATCAGTTTCTCTCTGTTGCAAGTGTGCACATCTCCATGGCATCTTCTGGCAA 1620
DB 1667 AGCCTCACATCAGTTTCTCTCTGTTGCAAGTGTGCACATCTCCATGGCATCTTCTGGCAA 1726
QY 1621 TACTGGGAACACAGCTTTTCCAGCAGCTCTCTCAGTGCCCTGCAAGCCATCAGTGAAG 1680
DB 1727 TACTGGGAACACAGCTTTTCCAGCAGCTCTCTCAGTGCCCTGCAAGCCATCAGTGAAG 1786
QY 1681 TCTGGGAGCTTCCCTTTTATCTACTCTGTATCACCAGGCCCCCAATGGATACTCTCC 1740
DB 1787 TCTGGGAGCTTCCCTTTTATCTACTCTGTATCACCAGGCCCCCAATGGATACTCTCC 1846
QY 1741 CAATATGAATATTACCCAAACCAAGTAAAGTAAGCAATCAGGATTCAGAGTCTCTCTGG 1800
DB 1847 CAATATGAATATTACCCAAACCAAGTAAAGTAAGCAATCAGGATTCAGAGTCTCTCTGG 1906
QY 1801 CTTTATTGCGAACCAATCCAGTGGAGTTCAATGTGTAGTCAATAGCAGAGATCA 1860
DB 1907 CTTTATTGCGAACCAATCCAGTGGAGTTCAATGTGTAGTCAATAGCAGAGATCA 1966
QY 1861 CCTCAGTGCAAGAAAGTAAGGAGGAGTGTCTCAGGGGCGAGAGATCAAGGGGTCC 1920
DB 1967 CCTCAGTGCAAGAAAGTAAGGAGGAGTGTCTCAGGGGCGAGAGATCAAGGGGTCC 2026
QY 1921 TTTGGAAGCAAGGTTCATAAAAAATTACTGCAGTTACTTACTCTGTCTCTGTATGACCG 1980
DB 2027 TTTGGAAGCAAGGTTCATAAAAAATTACTGCAGTTACTTACTCTGTCTCTGTATGACCG 2086
QY 1981 GGGTCAATCTCTCTTGACCAATCCCCCTAGATTCAAGTTGTAAGAAATCTTCTGTAG 2040
DB 2087 GGGTCAATCTCTCTTGACCAATCCCCCTAGATTCAAGTTGTAAGAAATCTTCTGTAG 2146
QY 2041 TCTCACCAGCCCTCTGGAGTCTCTCTCTACATCTGGAGAGTATCTCTACATCCAA 2100
DB 2147 TCTCACCAGCCCTCTGGAGTCTCTCTCTACATCTGGAGAGTATCTCTACATCCAA 2206

QY 2101 TATGATGGGTCACTGTTTACAGAGAAAGCACCGGATTTTGCAAGTTGCTGCAGAAATGG 2160
DB 2207 TATGATGGGTCACTGTTTACAGAGAAAGCACCGGATTTTGCAAGTTGCTGCAGAAATGG 2266
QY 2161 GAAATTCACAGCTGAGGTAGCCAAAGATTACTGCACAGCCACTGGGAAAGACACACAGAG 2220
DB 2267 GAAATTCACAGCTGAGGTAGCCAAAGATTACTGCACAGCCACTGGGAAAGACACACAGAG 2326
QY 2221 TATAACTTCTTGTGGGACGGAATGTTGTCAGCAGGAGCAGCTAAGTCTCTAGAGAA 2280
DB 2327 TATAACTTCTTGTGGGACGGAATGTTGTCAGCAGGAGCAGCTAAGTCTCTAGAGAA 2386
QY 2281 GGAGAAATTAATGCACTTCTTAGATACCTGCTGCACAGGATGATCCTTAGTGTGACTCTC 2340
DB 2387 GGAGAAATTAATGCACTTCTTAGATACCTGCTGCACAGGATGATCCTTAGTGTGACTCTC 2446
QY 2341 TAAAGAACTACAGCCCCCAAGTGGAGAGTGGACAAATAAATAGTCACTAGTCCACAGCTC 2400
DB 2447 TAAAGAACTACAGCCCCCAAGTGGAGAGTGGATAAATAAATAGTCACTAGTCCACAGCTC 2506
QY 2401 CACCAATCTAGCTCAAGTCAAGAGAAAGACCCCTAAATTAAGACAGAGACAAGTGAAGA 2460
DB 2507 CACCAATCTAGCTCAAGTCAAGAGAAAGACCCCTAAATTAAGACAGAGACAAGTGAAGA 2566
QY 2461 GGGATCTGGAGACTTGGATAATCTAGATGCTTATCTTGGTGTGATCTGACTAGTCTGACTT 2520
DB 2567 GGGATCTGGAGACTTGGATAATCTAGATGCTTATCTTGGTGTGATCTGACTAGTCTGACTT 2626
QY 2521 TTACAATAATTTCCATATCTCTCAATGGTAGTCTCTGGGGACTAAGCAACAGGTGTTTCA 2580
DB 2627 TTACAATAATTTCCATATCTCTCAATGGTAGTCTCTGGGGACTAAGCAACAGGTGTTTCA 2686
QY 2581 AGGAATTAATTTCTGCTGGTGTGAAAGTTTCACTGCTGCTGCTATCTCTCTCCATA 2640
DB 2687 AGGAATTAATTTCTGCTGGTGTGAAAGTTTCACTGCTGCTGCTATCTCTCTCCATA 2746
QY 2641 TAAACGAGCAGTGTCTCTGGATAGCCTCTTCTTGTGGTCAAGTCTCTCCAGTAAAAAA 2700
DB 2747 TAAACGAGCAGTGTCTCTGGATAGCCTCTTCTTGTGGTCAAGTCTCTCCAGTAAAAAA 2806
QY 2701 TATCAGTGTCTTCCCATATCTCTCAAGAACCAACCCATGTTGGTGGGAATCCAGAAATGAT 2760
DB 2807 TATCAGTGTCTTCCCATATCTCTCAAGAACCAACCCATGTTGGTGGGAATCCAGAAATGAT 2866
QY 2761 GGATAGTCAGAAATATATGCTCAAGTATGGGTGGGCCAAACCAATGCTGCTGTGAC 2820
DB 2867 GGATAGTCAGAAATATATGCTCAAGTATGGGTGGGCCAAACCAATGCTGCTGTGAC 2926
QY 2821 TCAGACTCTTCTCTCAGGAGACTGGGCTTACCAAACTCAAAAGCCGCGAGAATGGAAAC 2880
DB 2927 TCAGACTCTTCTCTCAGGAGACTGGGCTTACCAAACTCAAAAGCCGCGAGAATGGAAAC 2986
QY 2881 TATGAATTCAAACTTCATGGGAAGCAGGAGAGATTAATACTTTTACCAGACC 2940
DB 2987 TATGAATTCAAACTTCATGGGAAGCAGGAGAGATTAATACTTTTACCAGACC 3046
QY 2941 TGCAGTGGGTGGCTTATTTCCCATATTTCCCATATTTCCCATATTTCCCATATTTCCCATAT 3000
DB 3047 TGCAGTGGGTGGCTTATTTCCCATATTTCCCATATTTCCCATATTTCCCATATTTCCCATAT 3106
QY 3001 ACCAGTATTTGCAACAGCAGCAGCAGAGATGCTTCAATGAGGCTTGTGAAATCCCCCATGG 3060
DB 3107 ACCAGTATTTGCAACAGCAGCAGCAGAGATGCTTCAATGAGGCTTGTGAAATCCCCCATGG 3166
QY 3061 AATGGGGCTAATCCCTATGGCCCAAGCAGAGATCTAACCACTGGGTTCCTGGCCCGA 3120
DB 3167 AATGGGGCTAATCCCTATGGCCCAAGCAGAGATCTAACCACTGGGTTCCTGGCCCGA 3226
QY 3121 TGGCATCTGTGCTGAGGAGTCTCTCATGGCAGCTCAAAATAGGCTCTTCTTAGGAA 3180
DB 3227 TGGCATCTGTGCTGAGGAGTCTCTCATGGCAGCTCAAAATAGGCTCTTCTTAGGAA 3286
QY 3181 TTCCTTGGATGATCTTGTGGGCCACCTTCCCAACCTTGAAGGCCAGAGTGTGACAAAGAGC 3240

VQGIIRPPMAGCNSSVGGMSMSPNQLQMPSSRAYGLADPSTTGMGSGARXGSSNIA
SLTPGPMQSPSSQNNYRLNMSPLPHGSPGLAPNQNMISPNRNSPKTASHQFS
PVAGVHSPMSSGNTNFDSSSLALQAISEGVGTSLSTLSSPGPKLNDSPNMNI
TOPKSVNADSKSPGLFGYDQNPVSSMCSQNSRDHLSDEKSKSSVGEAENQRPLE
SGHKLLQDLCSDDRHSSLTNSPLDSSCKSSVSPTSPPGVSSSTGGVSSISN
MHGSLQEKHRLHLKLLQNGNSPAVAKITAEATGKDTSSITSCDGNVYKQELSPK
KKENALLRYLLDRDDPDLALKELOPQVGVDRKMSQCTSTIPSSQEKDPKIKTE
TSEEGSDNLDLDAILGDLTSSDFYNNISISSNGSHLTKQOVFOGTSLGKSSOSVQ
STRPYNLRAVNSPVSVCSSPPVKNISAFPMKPKPMLGNGNPRMDSQENYSSMGV
WGLPNKSGPEPMNSMGRPGDYNTSLRPPALGSGSIPILPLNSNIPGARNPVLOQ
QQQMLQMRPELPMGNGNPIYQAAASNLQNSWPDGMSLMEQVSHGTNRLPLNSLD
DLVGPSSNLEGGESDRALDQJHTLLSNTDGTGEEIDRALGIPELVNOGQALEPKD
AFQGEAAVMDQKAGLQYTPAQCPMPOGQFLQGGSPSFSNMMNMQNOGPNPLQ
GHPRANTMRPTNPTKOLRMOLOQLQOFLNGSRQALELKMENPTAGGAANVRPM
MOQOQFLNAOVNARSRELLSHFRQORVAMMOQOQOQOQOQOQOQOQOQOQOQ
OQOAFSPPNVTAAPSMDGLLAGPTMQAPQOFPYOPNYNGMGQOPQAPRGRVSSPP
NAMMSRMPGNPNMQRHQAASIIQSSEMKGPSGNLARNSSFSQOQFHLQKSPAVY
SVHNGSSGHGMQNMNPMPSGMPDPQKYC"

BASE COUNT 1955 a 1463 c 1511 g 1825 t

Query Match 96.1%; Score 4322.2; DB 11; Length 6754;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 4438; Conservative 0; Mismatches 8; Indels 81; Gaps 3;

QY	1	GCTGGATGGTGGACTCAGAGCAATAAATAAAGTCTTGAACATCTTTGACCTGGTT	60
DB	99	GCTGGATGGTGGACTCAGAGCAATAAATAAAGTCTTGAACATCTTTGACCTGGTT	158
QY	61	AGCCAGTGTGTATATTCAGATGAGTGGATTAGGAGAAAACCTTGGATCCACTGGC	120
DB	159	AGCCAGTGTGTATATTCAGATGAGTGGATTAGGAGAAAACCTTGGATCCACTGGC	218
QY	121	CAGTGATTCAGAAAACGCAATTTGCCATGTGATATCCAGGACAAGTCTTACCTGCAG	180
DB	219	CAGTGATTCAGAAAACGCAATTTGCCATGTGATATCCAGGACAAGTCTTACCTGCAG	278
QY	181	TGTTGAAAACGAGAGCGGAGCAGGAGAAAGTAAATATTTGAAGAATTTGCTGAGCTGAT	240
DB	279	TGTTGAAAACGAGAGCGGAGCAGGAGAAAGTAAATATTTGAAGAATTTGCTGAGCTGAT	338
QY	241	ATCTGCCAATCTTAGTGATATTGACAAATTTCAATGTCAACACCAAGATAATTTCCAAATGA	300
DB	339	ATCTGCCAATCTTAGTGATATTGACAAATTTCAATGTCAACACCAAGATAATTTCCAAATGA	398
QY	301	AAAGGAAACAGTAAAGCAGATACGTCAAATATAAGAGCAGGAGAAAACCTTATTTCCAAATGA	360
DB	399	AAAGGAAACAGTAAAGCAGATACGTCAAATATAAGAGCAGGAGAAAACCTTATTTCCAAATGA	458
QY	361	TGATGATGTTCAAAAACGCGATGTATCTTACAGGCGAGGAGTATTGATAAGACTC	420
DB	459	TGATGATGTTCAAAAACGCGATGTATCTTACAGGCGAGGAGTATTGATAAGACTC	518
QY	421	CTTAGGACCGCTTTACTTCAGGCAATTTGATGTTTCTTCTATTTTGGTGAATCGAGAGGC	480
DB	519	CTTAGGACCGCTTTACTTCAGGCAATTTGATGTTTCTTCTATTTTGGTGAATCGAGAGGC	578
QY	481	AAACATTTGATTTGATTCAGAAAATGTCACACAAATACCTGCAATATAAGCAGAGGACCT	540
DB	579	AAACATTTGATTTGATTCAGAAAATGTCACACAAATACCTGCAATATAAGCAGAGGACCT	638
QY	541	GGTTAACACAGGTGTTTACAAATATCTTACATGAAGAGCAGAAAAGGATTTCTTAAAGAA	600
DB	639	GGTTAACACAGGTGTTTACAAATATCTTACATGAAGAGCAGAAAAGGATTTCTTAAAGAA	698
QY	601	TTTACCAAAAATCTACAGTTTAAATGGAGTTTCTTGGACAAATGAGCCCAAGACAAAAAG	660
DB	699	TTTACCAAAAATCTACAGTTTAAATGGAGTTTCTTGGACAAATGAGCCCAAGACAAAAAG	758
QY	661	CCATACATTTAAATGCGGTATGTTGATGAAAACACCATGATATTCTTGGAGACATATAA	720
DB	759	CCATACATTTAAATGCGGTATGTTGATGAAAACACCATGATATTCTTGGAGACATATAA	818

QY	721	CGCCAGTCTCTGAAATGCGCCAGAGATATGAAACAATCGAGTCTTTGGCCTGTCTCAGCC	780
DB	819	CGCCAGTCTCTGAAATGCGCCAGAGATATGAAACAATCGAGTCTTTGGCCTGTCTCAGCC	878
QY	781	ACGAGCTATGATGGAGGAAGGGAAGATTGCAATCTTTGATGATCTGTGTGGCAGCGG	840
DB	879	ACGAGCTATGATGGAGGAAGGGAAGATTGCAATCTTTGATGATCTGTGTGGCAGCGG	938
QY	841	CATTACTACAGAGAAAGACATATTTCCATCAAAACCTTGAGAGCTTTATTACAGACATGA	900
DB	939	CATTACTACAGAGAAAGACATATTTCCATCAAAACCTTGAGAGCTTTATTACAGACATGA	998
QY	901	TCCTTCAGAAAAGTGTCAATATAGATACAAATTCACGTGAGATCCTCCATGAGGCTGG	960
DB	999	TCCTTCAGAAAAGTGTCAATATAGATACAAATTCACGTGAGATCCTCCATGAGGCTGG	1058
QY	961	CTTTGAAGATATAATCCGAAGGTGATTACAGAGATTTTATTAGTCTAAATGATGGGACGC	1020
DB	1059	CTTTGAAGATATAATCCGAAGGTGATTACAGAGATTTTATTAGTCTAAATGATGGGACGC	1118
QY	1021	ATGGTCCCAGAAAACGTCACTATCAAGA-----AGC	1050
DB	1119	ATGGTCCCAGAAAACGTCACTATCAAGAAGTTACCAAGTATGGGATATTTTCCCCAACAGC	1178
QY	1051	TTATCTTAATGGCCATGCAAAAACCCAGTATATCGATTCTCGTTGGCTGATGGAACTAT	1110
DB	1179	TTATCTTAATGGCCATGCAAAAACCCAGTATATCGATTCTCGTTGGCTGATGGAACTAT	1238
QY	1111	AGTGACTGCACAGACAAAAGCAAACTCTTCGGAATCCTGTAAACAAATGATCGACATGG	1170
DB	1239	AGTGACTGCACAGACAAAAGCAAACTCTTCGGAATCCTGTAAACAAATGATCGACATGG	1298
QY	1171	CTTTGCTCAACCCACTCTTCAGAGAGACAGAAATGGATATAGACCAAAACCCAAATCC	1230
DB	1299	CTTTGCTCAACCCACTCTTCAGAGAGACAGAAATGGATATAGACCAAAACCCAAATCC	1358
QY	1231	TGTTGACAAGGATTTAGACCACTTATGCTGGATGCAACAGTTTGGTAGGGGGATGAG	1290
DB	1359	TGTTGACAAGGATTTAGACCACTTATGCTGGATGCAACAGTTTGGTAGGGGGATGAG	1418
QY	1291	TATGTGCGCAAAACCAAGGCTTACAGATGCGAGCAGAGGGCTTATGGCTTGGCAGACCC	1350
DB	1419	TATGTGCGCAAAACCAAGGCTTACAGATGCGAGCAGAGGGCTTATGGCTTGGCAGACCC	1478
QY	1351	TAGCACCACAGGGCAGATGAGTGGAGTATGGGGTTCCAGTAAACATAGCTTCATT	1410
DB	1479	TAGCACCACAGGGCAGATGAGTGGAGTATGGGGTTCCAGTAAACATAGCTTCATT	1538
QY	1411	GACCCCTGGGCGCAGGATGCAATCACTTCTTACCAAGAACAACTATGGGCTCAA	1470
DB	1539	GACCCCTGGGCGCAGGATGCAATCACTTCTTACCAAGAACAACTATGGGCTCAA	1598
QY	1471	CATGAGTAGCCCCCAGATGGAGTCTTGGTCTTGGCCCAACACCAAGCAATATCATGAT	1530
DB	1599	CATGAGTAGCCCCCAGATGGAGTCTTGGTCTTGGCCCAACACCAAGCAATATCATGAT	1658
QY	1531	TTCTCGTCTGTAATGTTGGAGTCCAAAGATAGCTTCACATCAGTTTCTCTCTTGGCAGG	1590
DB	1659	TTCTCGTCTGTAATGTTGGAGTCCAAAGATAGCTTCACATCAGTTTCTCTCTTGGCAGG	1718
QY	1591	TGTGACTCTCCATGGGATCTTCTGGCAATACTTGGGAACCCAGAGCTTTCCAGAGAGCTC	1650
DB	1719	TGTGACTCTCCATGGGATCTTCTGGCAATACTTGGGAACCCAGAGCTTTCCAGAGAGCTC	1778
QY	1651	TCTAGTGCCCTGCAAGCCATCAGTGAAGGTGTGGGACTTCCCTTTTATCTACTCTGTC	1710
DB	1779	TCTAGTGCCCTGCAAGCCATCAGTGAAGGTGTGGGACTTCCCTTTTATCTACTCTGTC	1838
QY	1711	ATCACCAGGCCCCAAATGATGATACTCTCCCAATATGAATATTACCAACCAAGTAAAGT	1770
DB	1839	ATCACCAGGCCCCAAATGATGATACTCTCCCAATATGAATATTACCAACCAAGTAAAGT	1898

Qy	1771	AAGCAATCAGGATTC	CAAGAGTCCTCTGGGCTTTTATTGCGACCAAAATCCAGTGGAGAG	1833
Db	1899	AAGCAATCAGGATTC	CAAGAGTCCTCTGGGCTTTTATTGCGACCAAAATCCAGTGGAGAG	1958
Qy	1831	TTCAATCTGTGTCAGTCAAAATAGCAGAGATCACCTCAGTCGCAAAAGAAAGTAGGAGAGCAG	1890	
Db	1959	TTCAATCTGTGTCAGTCAAAATAGCAGAGATCACCTCAGTCGCAAAAGAAAGTAGGAGAGCAG	2018	
Qy	1891	TGTTGAGGGGCGAGAGAAATCAAAAGGGGTCCCTTTGGAAAGCAAAAGTGCATAAAAAATTACT	1950	
Db	2019	TGTTGAGGGGCGAGAGAAATCAAAAGGGGTCCCTTTGGAAAGCAAAAGTGCATAAAAAATTACT	2078	
Qy	1951	GCAGTTACTTACTCTGTCTTCTGTAGACCGGGGTCAATTCCTCTGTGACCAACTCCCCCT	2010	
Db	2079	GCAGTTACTTACTCTGTCTTCTGTAGACCGGGGTCAATTCCTCTGTGACCAACTCCCCCT	2138	
Qy	2011	AGATTCAAGTTGTAAAGAAATCTCTGTTAGTGTCAACGAGCCCTCTGGAGTCTCCTCCTC	2070	
Qy	2139	AGATTCAAGTTGTAAAGAAATCTCTGTTAGTGTCAACGAGCCCTCTGGAGTCTCCTCCTC	2198	
Qy	2071	TACATCTGGAGGAGTATCCTCTACATCCCAATATGCATGGGTCACTGTTTACAAGAAAGCA	2130	
Db	2199	TACATCTGGAGGAGTATCCTCTACATCCCAATATGCATGGGTCACTGTTTACAAGAAAGCA	2258	
Qy	2131	CCGATTTTGCACAAGTTGCTGCAGAAATGGGAAATTCACGACTGAGGTAGCGAAGATTAC	2190	
Db	2259	CCGATTTTGCACAAGTTGCTGCAGAAATGGGAAATTCACGACTGAGGTAGCGAAGATTAC	2318	
Qy	2191	TGCACAGCCACTGGGAAAGACACACGAGTATATACCTTCTTTGGGACCGGAATGTTGT	2250	
Db	2319	TGCAGAGCCACTGGGAAAGACACACGAGTATATACCTTCTTTGGGACCGGAATGTTGT	2378	
Qy	2251	CAACGAGGACGACTAAGTCTCTAAGAAAGAGAGAAATATGCACCTTCTTAGTACCTGCT	2310	
Db	2379	CAACGAGGACGACTAAGTCTCTAAGAAAGAGAGAAATATGCACCTTCTTAGTACCTGCT	2438	
Qy	2311	GGACAGGGATGATCCTTAGTGATGCACCTCTCTAAAGAACTACAGCCCCAAGTGGAGGAGT	2370	
Db	2439	GGACAGGGATGATCCTTAGTGATGCACCTCTCTAAAGAACTACAGCCCCAAGTGGAGGAGT	2498	
Qy	2371	GGACAATAAANTGAGTCAGTCGACACGCTCCACCATTCTTAGCTCAAGTCAAGAAAGA	2430	
Db	2499	GGATAATAAANTGAGTCAGTCGACACGCTCCACCATTCTTAGCTCAAGTCAAGAAAGA	2558	
Qy	2431	CCCTAAAAATTAAGACAGACAAAGTGAAGAGGGATCTGGAGACTTGGATATCTAGATGC	2490	
Qy	2559	CCCTAAAAATTAAGACAGACAAAGTGAAGAGGGATCTGGAGACTTGGATATCTAGATGC	2618	
Qy	2491	TATCTTTGGTGATCTGACTAGTTCTTGACFTTTTACAATAATTCATATCTCCAAATGGTAG	2550	
Db	2619	TATCTTTGGTGATCTGACTAGTTCTTGACFTTTTACAATAATTCATATCTCCAAATGGTAG	2678	
Qy	2551	TCATCTGGGACATAAGCAACAGGTGTTTCAAGGAACATAATTCTCTGGGTTTGAAGTTTC	2610	
Db	2679	TCATCTGGGACATAAGCAACAGGTGTTTCAAGGAACATAATTCTCTGGGTTTGAAGTTTC	2738	
Qy	2611	ACAGTCTGTGCAGTCTATTCTGCTCCATATTAACCGAGCAGTGCTCTGAGATAGCCCTGT	2670	
Db	2739	ACAGTCTGTGCAGTCTATTCTGCTCCATATTAACCGAGCAGTGCTCTGAGATAGCCCTGT	2798	
Qy	2671	TTCTGTTGGTCAAGTCTCCAGTAAAAAATATCAGTGCCTTTCCCATGTTTACCAAAAGCA	2730	
Db	2799	TTCTGTTGGTCAAGTCTCCAGTAAAAAATATCAGTGCCTTTCCCATGTTTACCAAAAGCA	2858	
Qy	2731	ACCCATGTTGGTGGGAATCCAGAAATGATGATAGTCAGGAAAAATTATGGCTCAAGTAT	2790	
Db	2859	ACCCATGTTGGTGGGAATCCAGAAATGATGATAGTCAGGAAAAATTATGGCTCAAGTAT	2918	
Qy	2791	GGGTGGGCCAAACCGAAATGTGACTGTGACTCAGACTCCTCTCCCTCAGAGACTGGGCTT	2850	
Db	2919	G-----GGAGACTGGGCTT 2933		
Qy	2851	ACCAAACTCAAAGCCCGGAGAAATGGAACTATGAATTCAAACTCCATGGGACAGCCAGG	2910	

[illegible]

Db 445 CTCCTTAGGACCGCTTTTACTACAGGCACTGGATGGTTTCTGTTTGGTGAATCGAGA 504
Qy 478 GGCAAAACATTGTATTTGTATCAGAAAATGTACACAAATACCTGCAATATAAGCAAGAGGA 537
Db 505 TGGAAACATTGTATTCGTGTGTCAGAAAATGTACACAGTATCTGCAGTACAAGCAGAGGA 564
Qy 538 CCTGGTTAAACACAGTGTTCACATATCTTACATGAAGAAGACAGAAAGATTTCITAA 597
Db 565 CCTGGTTAAACACAGTGTCTACAGCATCTTACATGAGCCAAAGCAGGAAGGA-TTCTTTAA 623
Qy 598 GAATTTACCAAAATCTACAGTTAATGGAGTTTCTGGCAAAATGAGCCCCCAAGACAAA 657
Db 624 ACACCTTACCAAAATCCACAGTTAATGGAGTTCTTGGACTAATGAGAACAGACAAA 683
Qy 658 AAGCCATACATTTAATTCGCGTATGT-TGATGAANAACACCATGATGATTTCTGGAAAGACA 716
Db 684 AAGCCCCATACATTTAATTTGCTGCTATGTTGATGTAAGAAACACACGACATTTTGAAGAGC 743
717 TAAAGCCAGTCCCTGAATGCGCCAGACATATGAACAATCAGTGTCTGCCCTGTCTC 776
744 TGAATGCCAGTCCCGAAACGCCAGAGATATGAACAATCAGTGTCTGCCCTGTCTC 803
Qy 777 AGCCACGAGCTATGATGGAGGAAGGAGATTTGCAATCTGTATGATCTGTGTGGCAC 836
Db 804 AGCCTCGCGCTATGCTGGAAGAGAGAGACTTGCAGTGTGTATGATCTGCTGGTGC 863
Qy 837 GCGCGATTACTACAGGAGAGAAACATTTCCATCAAAACCTCGAGAGCTTTTATTACCAGAC 896
Db 864 GCGCGTGACTGC-----GCCATCCCATCCAGTCCCGAGAGCTTTATTACCAGAC 914
Qy 897 ATGATCTTTACGAAAGTGTGCAATATAGATACAAATCTACTGAGATCTCCATGAGGC 956
Db 915 ATGACCTTTCCGGAAGTGTGCAATATAGATACAAATCTACTGAGTCTTCCATGAGGC 974
Qy 957 CTGGCTTTGAAGATATAATCCGAAGTGTATTACAGAGATTTTATAGTCTAAATGATGGC 1016
Db 975 CTGGCTTTGAAGACATATCCGAAGATGTATCCAGAGGTTCTTCACTGATGATGGGC 1034
Qy 1017 AGTCATGTGCCAGAAAGCTCACTATCAAGAAGCTTATCTTAATGGCCATCAGAAACCC 1076
Db 1035 AGTCATGTGCCAGAAAGCTCACTATCAAGAAGCTTATGTTTCATGGCCACGACAGACCC 1094
Qy 1077 CAGTATATCGAATCTCGTTGGCTGATGAACTATAGTACTGCACAGACAAAAGCAAC 1136
Db 1095 CCGTGTATCGTTTCTCTTGGCTGATGAACTATGTGAGTGGCGACACAAAAGCAAC 1154
1137 TCTTCCGAAATCCTGTAACAAATGATCCACATGGCTTTGCTCAACCCACTTCTTCAGA 1196
1155 TCTTCCGCAATCTGTAAAGATGATCTCAGGGTTTCACTCGACCCACTTCTTCAGA 1214
Qy 1197 GAGAACAGAAATGGATATAGACCAACCCAAATCTGTTGGACAGGGATTAAGACCCCTA 1256
Db 1215 GAGAACAGAAATGGATATAGACCAACCCAAATCTGTTGGACAGGGATTAAGACCCCTG 1274
Qy 1257 TGGCTGGATGCAACAGTTCCGTTAGCGGCGATGAGTATGTGCGCAACCAAGCTTTACAGA 1316
Db 1275 CAGCAGGCT-----GTGGCGTGAGCATGTCTCCAAATCAGAATGTACAGA 1319
Qy 1317 TGCCGAGCAGCAGGCGCTATGCTTGGCAGACCCCTAGCACACAGGGCAGATGAGTGAG 1376
Db 1320 TGATGGGAGCGCGACCTATGCGTGGCAGACCCCTAGCACACAGGGCAGATGAGTGAG 1379
Qy 1377 CTAGTATGGGGTTTCCAGTAACATAGCTTTCATGACCCCTGGGCGACGCGATGCAATCAC 1436
Db 1380 CTAGTACGGGGCTTCTAGTAGGCTAGCTCACTGACGCCAGGACAAAGCCTACAGTCG 1439
Qy 1437 CATCTTCCTACAGAACCAACACTATGGGCTCAACATGAGTAGCCCCCAGCATGGAGTC 1496
Db 1440 CATCTTCCTATCAGAACAGCAGCTATGGCTCAGCATGAGCAGTCCCCCAGCAGTC 1499
Qy 1497 CTGGTCTTGGCCCCAACACAGCAGATATCATGATTTCTCTCGTAAATCGTGGGAGTCCAA 1556
Db 1500 CTGGTCTTGGCCCCAACACAGCAGAAACATCATGATTTCTCCCTCGGAATCTGTGGCAGCCAA 1559

Qy 1557 AGATAGCCTCACATCAGTTTTTCTCTGTGCGAGTGTGCACACTCTCCCATGCATCTTCTG 1616
Db 1560 AGATGGCTCCACCAGTCTCTCTGTGCGAGTGTGCACACTCACCCATGGACCTTCTG 1619
Qy 1617 GCAATACTGGGAACACAGCTTTTCCAGCAGCTCTCTCAGTGGCCCTGCAAGCCATCAGT 1676
Db 1620 GCAACAGGAGGACACAGCTTTTCTAGCAGTCCCTCAGTGGCTTGCAGCCATCAGT 1679
Qy 1677 AAGTGTGGGACCTTCCCTTTTATCTGTCTCATCCACAGGCCCCAAATTTGGATAACT 1736
Db 1680 AAGCGTGGGACCTCTCTTTTATCTGTCTCTCACCAGGCCCCAAATTTGGATAAT 1739
Qy 1737 CTCCCAATATGAATATACCCAAACAAGTAAGTAAGCAATCAGGATTTCCAAAGTCTC 1796
Db 1740 CTCCCAATATGAATATAGCCAGCCAAAGTAAGTGAAGTGGTTCAGGACTCTAAGAGCCCC 1799
Qy 1797 TGGCTTTTATTTGGACCAAAATCCAGTGGAGATTCANTGTCTCAGTCAATAGCAGAG 1856
Db 1800 TAGGCTTATCTGTGAACAGAAATCCAGTGGAGATTCAGTGTCTCAGTCAACACAGCAG 1859
Qy 1857 ATCACCTCAGTGACAAAGAAAGTAAGGAGACGAGTGTGAGGGGCGACAGAAATCAAAGG 1916
Db 1860 ATCCCAAGTGAANAAGAAAGAGAGAGCAGTGGGAGGTGTGAGAGAGCCAGG 1919
Qy 1917 GTCTTTGGAAAGCAAGGTCTATAAAATTTACTGCACTTACTTACCTGTTCTTCTGATG 1976
Db 1920 GACCTCTGGAAGCAAGAGCCCAAGAAATGCTGCACTTACTCAGTGTCTCTCCGAGC 1979
Qy 1977 ACCGGGTCTATCTCTTGACCAACTCCCCCTTAGATTCAAGTTGTAAAGATCTTCTG 2036
Db 1980 ACCGAGGCCATCTCTCTTGACCAACTCTCCCTGGATTCACAACTGCAAGAGCTTCTCG 2039
Qy 2037 TTAGTGTCAACAGCCCTCTGAGTCTCTCTCTACATCTGGAGAGTATCTCTCTACAT 2096
Db 2040 TTAGTGTCAACAGCCCTCTGAGTGTCTCTCTCAACATCAGGAGACAGTGTCTCCACT 2099
Qy 2097 CCAATATGATGCTGCTACTGTTACAAGAAAGACCCGATTTTGCACAAAGTGTCTGAGA 2156
Db 2100 CCAATGTGATGCTGCTCTGTTGCAAGAGAAACACCGGATTTTGCACAAAGTGTCTGAGA 2159
Qy 2157 ATGGAATTCACAGCTGAGTACCAAGATTTACTGCACAAAGCCACTGGGAAGACACCA 2216
Db 2160 ATGCAACTCTCCAGCGGAGTCTGCAAGATCTGCAAGAGCCACTGGGAAGACACCA 2219
Qy 2217 GCAGTATAACTTCTGTTGGGACGAAATGTTGCAACGAGGAGCAGCTAAGTCTCTAAGA 2276
Db 2220 GCAGACTGCTTCTGTTGGAGAGGGGA---CAACAGCAGCAGGAGCAGTGTCTAAGA 2276
Qy 2277 AGAAGGAAATATGCACTTCTTAGATACCTGCTGGACAGGGATGATCTTAGTGTGAC 2336
Db 2277 AGAAGGAAATATGCTCTGCTTAGATACCTGCTGGACAGGGATGATGATGATGTC 2336
Qy 2337 TCTCTAAGAACTACAGCCCCCAAGTGGAGAGTGGACAAATAAATGAGTGTGAGTGTGACCA 2396
Db 2337 TTGCAAAAGAGCTGCAAGCCCCCAGCAGTGGGAGACAGTAAATGAGTGTGAGTGTGAC 2396
Qy 2397 GCTCCACCATCTCTAGCTCAAGTCAAGAGAAAGACCCCTAAATTAAGACAGACAGCAAGT 2456
Db 2397 GCTCCACCATCTCCAGCTCTGCGCAAGAGAAAGACCCCAAAATTAAGACAGGAGCAAG 2456
Qy 2457 AAGAGGATCTGGAGACTTGGATATCTAGATGCTATTTCTTGGTGTGATCTGACTAGTCTG 2516
Db 2457 ACAGGATCTGGAGACTTGGATATCTAGATGCTATTTCTTGGAGATTTGACAGTGTG 2516
Qy 2517 ACTTTTACAATAATTCATATCTCAATGTAGTGTATCTGTGGGAGTGTGAGCAACAGTGT 2576
Db 2517 ACTTCTACAATAATCTA-----CAATGGCGGTTCACCCAGGCGCCAAACAGCAGATGT 2570
Qy 2577 TTCAGGAACTAATCTCTGGGTTTGAAGATTCACAGTGTGCACTGTGCTGCTTCTGCTC 2636
Db 2571 TTGAGGAGCAGGTTCTCTGGGTTTGGGAAGTCCACAGCCTGTGCACTGTGCTGCTC 2630

Qy	2637	CATATACCGAGCAGTGTCTCTGGATAGCCCTGTCTTCTGTGGCTCAAGTCTCCACAGTAA	2696
Db	2631	CATATACCGAGCGGTGTCTCTGGATAGCCCTGTCTCTGTGTGGCTCAGGTCCGCGAGTGA	2690
Qy	2697	AAATATACAGTGTCTTCCCATGTTACCAAAACCAACCATCTGTGGGTGGGAATCCAGAA	2756
Db	2691	AGAAATGTCTAGTGTCTTCCCTGGGTTCACAAAACAGCCCATCTAGCTGGGTGGGAATCCAGAA	2750
Qy	2757	TGATGATAGTCAGGAAAAATATATGGCTCAAGTATGGGTGGGCCCAAAACCGAAATGTGACTG	2816
Db	2751	TGATGATAGTCAGGAGAAATACGGTGGCCACATG---GGCCAAAACAGAAATGTTCTCTG	2807
Qy	2817	TGACTCAGACTCTCTCTCTCAGAGACTGGGGCTTACCAAACTCAAAAGCCCGCAGAAATGG	2876
Db	2808	TGAATCCGACTTCTCTCCCGGAGACTGGGGCTTAGCTAACTCAAGGGCCAGCAGAAATGG	2867
Qy	2877	AACTTATGAATCAAACTCCATGGGAAGACCAGGAGGAGATTAATAATCTTCTTACCCTA	2936
Db	2868	AGCTCTGGCATCAAGTCCCTCGGGAAGAACTTGGAGCCGATTAACAGTCCCATTTACCCTA	2922
Qy	2937	GACTTGCACTGGGTGGCTCTATTCCACATTTGCCCTTCCTGGCTCTAATAGCATACACAGTGTG	2996
Db	2928	GACTTGCCATGGGGGCTCTGTGCTACTCTTGCCACTTCGTCTTAATCGACTGCCAGTGTG	2987
Qy	2997	CGAGACCAGTATT-----	3010
Db	2988	CAAGACATCGTTTGACAGCAACAGCAGCAGCAACAGCAGCAACAGCAACAGCAGCAGC	3047
Qy	3010	-----GCNACAGCAGCAGCAGATGCTTCAATGAGGCTTGGTGAATCCCNATGGGAA	3062
Db	3048	AACAGCAGCAGCAACAGCAGCAGATGCTTCAATAGAGAACTGGTGAGATTCCTCATGGGAA	3107
Qy	3063	TGGGGCTTAATCCTATGTGCCAAGCAGCAGCATCTFAACCAACTGGTTCCTGGCCCGATG	3122
Db	3108	TGGAGTCAATCCNTATACCCACAGCAGTGCAGTCTAACCACACAGGTTCTTGCCAGAGG	3167
Qy	3123	GCATGTTGCCATGGAAAGTTTCTCATGGCACTCAAAATAGGCCCTCTTCTTAGGAATT	3182
Db	3168	GCATGCTCTCATGGAACAAGGCTCTCAGGGTCTCAAAATAGGCTCTTCTTAGAACT	3227
Qy	3183	CCCTGATGATCTGTTGGGCCACCTTCCAACTGGGAAGGCCAGAGTGACGAAGAAGCAT	3242
Db	3228	CTCTGATGATCTGCTTGGGCCACCTTCTAACGCAGAGGGCCAGAGTGACGAGAGCTC	3287
Qy	3243	TATTGACAGCTGCACACTCTCTCAGCAACACAGATGCCACAGGCTCGGAGAAATTG	3302
Db	3288	TGCTGGACAGCTTGACACATCTCTGAGCAACACAGATGCCACAGGTTCTGGAGGAGATCG	3347
Qy	3303	ACAGAGCTTTGGGCATTCTGAACTTGTCAATCAGGGACAGGCATTAGAGCCCAACAGG	3362
Db	3348	ACAGGGCTTTGGGAATCTTGAGCTCGTGAATCAGGGACAGCTTTGGAGTCCAAACAGG	3407
Qy	3363	ATGCTTTCCAAAGCCCAAGACGAGCAGTAATGATGGATCAGAAGCAGGATATATGGAC	3422
Db	3408	ATGTTTTCCAAAGCCCAAGACGAGCAGTAATGATGGATCAGAAGGCTGCATATATGGAC	3467
Qy	3423	AGACATACCCAGCAGAGGGCTTCCAAATGCAAGGAGGCTTTCTATCTTCAGGAGCAATCAC	3482
Db	3468	AGACATACCCAGCTAGAGGTCTCCCTCTCAAGGAGGCTTTAACCTTCAGGAGCAGTCA	3527
Qy	3483	CATCTTTAACTCTATGATGAATCAGATGAACACGACAGGCAATTTTCCCTCTCCAAGGAA	3542
Db	3528	CATCGTTTAACTCTATGATGGGTCAAGTTAGCCACCAAGGCAGCTTTCTCTTCGACGGCA	3587
Qy	3543	TGCACCCACGAGCAACATCATGAGACCCCGGACAAACACCCCAAGCAACTTAGAATGC	3602
Db	3588	TGCATCTAGAGCCGCCCTCGTGAGACCAAGSACCAACACCCCGAAGCAGCTGAGAAATGC	3647
Qy	3603	AGCTTCAGCAGAGGCTGCAGGGCCAGCAGTCTTTGAAATCAGAGCCGACAGGCACCTGAAT	3662
Db	3648	AGCTTCAGCAGAGGCTACAGGGCCAGCAGTCTTTTAAATCAGAGCCGCGACACTTGAA	3707
Qy	3663	TGAAATGGAACCCCTACTGCTGTGTGGTCTGCGGCTGATGAGGCTCTATGATGACGCCCC	3722

Db	3708	TCGAAATGGAGAAC-----CTGCTGGCACTCTGTGATGAGGCCCATGATG-----C	3755
QY	3723	AGCAGGGTTTTCTTAATGCTCAAAATGTCGCCAAGCAGCAGCAGAGCTGCTTAAGTCATC	3782
Db	3756	CCCAGGCTTTCTTTATGCCCCAATGCTGCCCAGCAGAACAGAGCTGTATGAGCCATC	3815
QY	3783	ACTTCGACACAGAGAGGGTGGCTATGATGATGAGCAGCAGCAGCAACAGCAGCAGCAGC	3842
Db	3816	ACCTGCAGCAGCAGAGGATGGCGATGAT-----	3847
QY	3843	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCC	3902
Db	3847	-----GTACAAACCAACAGCCTCAGGCCCTTCAGCC	3875
QY	3903	CACCTCCTTAATGTGACTGCTTTCCGCCAGCATGGATGGCTTTTGGCAGAGACCCACAAATGC	3962
Db	3876	CACCTCCCAACGTCACCGCCTCCGCCAGCATGGACGGGTTTTTGGCAGGTTTCAGCAATGC	3935
QY	3963	CACAAGCTCTCCGCAACAGTTTCCATATACAAATATGAATGGGAGCAACACCAG	4022
Db	3936	CGCAAGCCCTCCACAAACAGTTTCCATATCAGCAAAATTACGGAATTGGCAACACCACAG	3995
QY	4023	ATCCAGGCTTTGGTCGAGTGTCAGTCTCCCAATGCAATGATGTCGTCGAAGAATGGGTC	4082
Db	3996	AGCCAGCCTTTGGTCGAGGCTCGAGTCTCCAGTGCATGATGTCATCAAGAATGGGGC	4055
QY	4083	COTCCAGAATCCCATGATGCACACCCGAGGCTGCATCCATCTATCAGTCTCCTCAGAAA	4142
Db	4056	CTTCCAGAATGCCATGGTGCAGCATCTCCAGCCACACCCATGTATCAGCCTTCAGATA	4115
QY	4143	TGAAGGCTGGCCATCAGGAATTTGGCCAGGACAGCTCCCTTTCCAGCAGCAGTTTG	4202
Db	4116	TGAAGGGTGGCCGTCAAGGAACCTGGCCAGGAATGGCTCTTCCCCCAAGCAGCAGTTTG	4175
QY	4203	CCCAAGGGGAATCCTCGCAGTGATAGTGGTGCACATGAATGGCAGCAGTGGTCACA	4262
Db	4176	CTCCCCAGGGAAACCTTCAGCCTTACAACATGGTGCATATGACAGCAGGGTGGCAGCT	4235
QY	4263	TGGGACAGATGAACATGAACCCCATGCCCATGTCGTGGCATGCCTATGGTGTCTGTACGA	4322
Db	4236	TGGGACAGATGGCATGACCCCATGCCCCATGTGTGGCATGCCATGGGCCCGCCGATCAGA	4295
QY	4323	ATACTGCTGACATCTCTGCAACGAGACCTCTTAAGGAACACCATGTACAAATGACACTG	4382
Db	4296	ATACTGCTGACATCTCCCTAGTGGGAC-----TGACTGTACAGATGACATGCGACAG	4348
QY	4383	CAC TAGGATTATTTGGGAAGGAATCATTTGTTCCAGGCATCCATCTTGGAGAAAGGAC	4439
Db	4349	GATCATCAGGACGTTGGCGGCAGTCATTGTCTAAGCATCCAGCTTGAAGCAAGGCC	4405

RESULT	6
AF044080	
LOCUS	Xenopus laevis retinoid X receptor-interacting coactivator xSRC-3
DEFINITION	mRNA, complete cds.
ACCESSION	AF044080
NID	G2852393
VERSION	AF044080.1 GI:2852393
KEYWORDS	African clawed frog.
SOURCE	Xenopus laevis
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
	Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae;
	Xenopus.
REFERENCE	1 (bases 1 to 4546)
AUTHORS	Khm,H.-J., Lee,S.-K., Na,S.-Y., Choi,H.-S. and Lee,J.W.
TITLE	Molecular cloning of xSRC-3, a novel retinoid X receptor-interacting coactivator from Xenopus, that is related to AIB1, p/CIP and Tir2
JOURNAL	Mol. Endocrinol. (1998) In press
REFERENCE	2 (bases 1 to 4546)

AUTHORS Kim, H.-J., Lee, S.-K., Na, S.-Y., Choi, H.-S. and Lee, J.W.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-1998) Pharmaceutical Sciences, Chonnam National Univ., 300 Yongdong-gong Puk-ku, Kwangju 500-757, Korea
FEATURES
source 1. .4546
/organism="Xenopus laevis"
/db_xref="taxon:8355"
CDS
1. .4176
/note="member of the SRC-1 superfamily of transcription coactivators"
/codon_start=1
/product="retinoid X receptor-interacting coactivator xSRC-3"
/protein_id="AAC12927.1"
/db_xref="PID:g2852394"
/db_xref="GI:2852394"
translation="MSGIGENSLDPLASETRKRKPPSSCDTPGGLTCSGKRRREQES
KYLEELADLISANLSDIDNENPDKCAILKEVYRQIQKEQGKASNNDDVQKADY
STGGQVIDKDSGLPLQALDGLFLYVNNREGSIVFVSENVTLQIKQEDLVNTSVI
SILHEEDRKDFLKNLPKSTVNGVPWFSTPRQKSHTFNCRMLVKTSHDLEDNSLDA
ROYETMQCFLSOPRAMIEGEDLQCMICVARRITTAERAFSANPESTIRHDTLY
KVNIDANLSRSMRPGFEDTIRRCIORFLFHSEGOPTVYKRYHOEAYVHGLSETPL
RESLADCTMYATQKSLFRNPVNDPHGFVSTHFLQREONGYRPNPNNAQGTROP
NPPLPNTMNSPQAMQOQNNGVNDPNSMAQOQRYKSPGNMAVYNAQVQOQSP
IQNNNTGLNNSPFGPSGMNANQPNLMSVPRNRAFSMAQOQSPVPMGNPMGSS
GNAGGFSFSSLSALHAISEGVSSLSLSLSPQKVENNSNMNPQOQKICNDCK
SPGLYCEQOQVSCQSGREHLGKEDKENIFEGESQROAESKQGHKLILQLIT
CFTEERGOSLSSMSMDCKDSNVTSVSGVSSSTSIGVSTSLNHLGSMLOEKHRLHK
LLONGNSPAEVAIKTAEATKDVFQETVSSAPCTEATVREQLSPKPKENALLRHL
DKDDKPLADKIDPKVEHMDIKMGSSSSNVTSVSSQDKVKIKTEREEVPGDNL
DAILDGLDFTSNMSRASDLGPKVPQDSPTLAMSPOSMQSGSRPPFNAMSL
DSRSSTPPNRVNSFPLPKQMGIDGSPMDGQDNFGVMGSPNRSNMQPFGDNAM
QNAVNLEPNVSGVPGPDYSAMTRAMPGNMPLLRNSIPGSRPVMOQOQH
ILPNRNDMAMSGSNPYGOQASNPSPWPDALMMNGRGQNRNOLGRNSLDLLC
PPSTVEGQDEIALLDOLHTLLSNLDTGLEEDRALGIPDLVSOGALPQDPSYOP
QGSVIDQRPYQHYAGQAMSGAGNNQOQHPPTVYNGVQOQGHPLPG
MHPNANLIRNNILKOLRMLQORLOQGLNORALEMKVDPMNPGGAGVPRVYM
QTVPSQOGLFNAQVAKQKRELISHQIRHMAQMMQOQOQPOAFSPPNVYASAM
DPLVGGPMPQAPQOQSPYNTYQINQOQDTPFGRVSSPPNAMSRSMAQSPNPHT
TOMYPSPDMKGPNSMARPNSPPOQYSHQTNPATYNNMHNMGNGHMQMINSLP
MSGMPMGPDQKCYC"
BASE COUNT 1309 a 1150 g 957 t
ORIGIN

Query Match 39.9%; Score 1793.4; DB 4; Length 4546;
Best Local Similarity 66.7%; Pred. No. 0;
Matches 2856; Conservative 0; Mismatches 1291; Indels 138; Gaps 15;

Qy 86 ATGAGTGGATTAGGAGAAAA---CTTGGATCCACTGGCCAGTGATTCACGAAAGCGCAA- 142
Dy 1 ATGAGTGGATTAGGAGAAAACTCTTGGATCCGCTGGCGTCTGAAACAAAGAGCGCAAG 60

Qy 142 --ATTGCCATGTATCTCAGGACAGAGTCTTACCTGCATGGTGGTGAAGAACGGAGAGCG 199
Dy 61 CCGTCTCTCTGTATACCCCGGACCTGGTCTGACCTGAGTGGTGAAGAGAGGAGCGGT 120

Qy 200 GAGCAGGAGAAATATATATTGAAGAAATGGCTGAGCTGATATCTGCCAAATCTTAGTGAT 259
Dy 121 GAGCAGGAGAGCAATGATAGAGAGTGGCCGATCTCATTTACGCCAATCTTAGCGCAC 180

Qy 260 ATTGACAATTTCAATGTCAAAACAGATAAATGTGCGATTTTAAAGGAAACAGTAAGACAG 319
Dy 181 ATAGACAATTTCAATGTCAAAACAGATAAATGTGCGATTTTAAAGGAGAGCGTACGGCAG 240

Qy 320 ATAGCTCAATAAAGACAGAGAAACTATTTCCATATGATGATGATGTTTCAAAAAGCC 379
Dy 241 ATCCCTCAGATTAAGAGAGAGAAAGCCCTCTTCTAATAGCAGGAGTGTGCAAGAAAGCC 300

Qy 380 GATGATCTCTACAGGCGGAGGAGTTATGATAAGAACTCTTAGGACCGCTTTTACTT 439
Dy 301 GATGTTCTCTACAGGCGGAGGAGTGTGATAAGAGGATCTCTCGGGCGCTGTATTG 360

Qy 440 CAGGCATTTGGATGGTTTCCCTATTTTGGTGAATCGAGAGCAAAACATTTGTTATGATCA 499
Dy 361 CAGGCTCTGGATGGCTTTCTGTATGTGAATCGAGAGGCGAGCATGTTATTTGTTATCT 420

Qy 500 GAAATGTCACACATACCTTCGAATATGAAGCAAGAGACCTGGTTACACAAGTGTTCAC 559
Dy 421 GAAATGTCACCCAGTACTTACAATAACAAGAGGATCTGGTTAAACACTAGTGTCTAC 480

Qy 560 AATATCTTACATGAAGAGACAGAAAGGATTTCTTAAAGAAATTTACCAAAATCTACAGTT 619
Dy 481 AGCATTTCGACGAGAGGACCGTAAGATTTCTCAAGAACCTGCCCAATCTACTGTT 540

Qy 620 AATGGAGTTTCTCGACAAATGAGCCCCAAAGACAAAGAACCCATACATTTAAATGGCGGT 679
Dy 541 AATGGCGTCCCATGGTTTAGTGAACCCCTCGACAGAAGAGCCACACGTTCAACTGCCGC 600

Qy 680 ATGTTGATGAAGAACACACATGATATTCTGGAAGACATAAAGCCAGTCTCTGAATGGCG 739
Dy 601 ATGCTGGTCAAAACCTCTCATGATCATCTGGAGGACGGAA---GCAACCTGGACCGCGCT 657

Qy 740 CAGAGATATGAACAAATGCGAGTGTCTTGGCCCTGCTCAGCCACGAGCTATGATGGAGAA 799
Dy 658 CAGAGATACGAGACCATGCGAGTGTCTTGTATCTCAGCCCCGGCGCATGATCGAGAG 717

Qy 800 GGGGAAGATTTGCAATCTTGTATGATCTGTGTGCGCCGCGCATTTACTTACAGGAGAAAG 859
Dy 718 GGAGAAGACCTTCAGTCTCTGTATGATCTGTGTGCGCCAGGCGCATCACAACTGCAGAGG 777

Qy 860 ACATTTCATCAAAACCTTGAGAGCTTTATACAGACATGATCTTTACGAGAAAGTGTGTC 919
Dy 778 GCCTTTCAGCCAACTCTGAAAGCTTCATCAAGGCGACGATCTGACCGGCAAGGTGGTG 837

Qy 920 AATATAGATCAAAATCTCACTGAGATCTCTCCATGAGGCTGTGCTTTGAAGATTAATCCGA 979
Dy 838 AACATTGATGCAACTCCCTCGCTCTCCATGAGAGCGGGCTTTGAAGATACCATCGT 897

Qy 980 AGGTGTTATCAGAGATTTTGTAGTCTAAATGATGGGAGTGTGTTCCAGGAAAGCTCAC 1039
Dy 898 AGATGTTATCAAGGTTCTCTGTTTACAGGCGAAGGCGAGCCATGACCTACAAAGCGCAT 957

Qy 1040 TATCAAGAGCTTATCTTAATGGCCATGCGAAACCCAGTATATCGATTTCTGTTGGCT 1099
Dy 958 TACCAAGAGGCTATGTCACGAGCTTTCTGAAACGCTCTGTATAGTCTCTCTAGCA 1017

Qy 1100 GATGGAATCTAGTACTGACACAGACAAAGCAAACTCTTCCGAAATCTCTGTAAACAAT 1159
Dy 1018 GATGGACCATGGTGTGCTCTCAGACAAAGTAAGTCTTCCGGAATCTCTGTACCCT 1077

Qy 1160 GATCGACATGGCTTGTCTCAACCCACTCTCTTCAGAGAGAACAGATGGATATAGACCA 1219
Dy 1078 GATCGGACGCTTTGTCTCCACGATTTCTTGCAAGGGAACAGATGGCTACCGACCC 1137

Qy 1220 AACCAATCTGTGTGACAGAGGATTTAGACCCTATGGCTGGATGCAACAGTTCGGTA 1279
Dy 1138 AATCCGAACCAATGGCTCAAGGTATTTCGGCCCAAAATG-----AACCCCTAACCTA 1188

Qy 1280 GGGGCGATGATGTGCGCCAAACCAAGGCTTACAGATCGCGAGCAGCGCCCTATGGC 1339
Dy 1189 CCCAATCAATGAATTAATGCCAGCGCCATGAGAGAGCAAGCAAGCAATATGGC 1248

Qy 1340 TTGGCAGACCTTAGCACACACAGGCGAGATGATGGAGCTAGGTATGGGGTTCCAGTAAC 1399
Dy 1249 ATGGGGACCCCAACAGCATGCGACAAATGCGAGGCGATGATACAACTCCCCAGGCAAC 1308

Qy 1400 ATAGCTTATTCACCCCTGGGCGAGGATGCAATCACCCTCTTCTACCGAAGCAACAAAC 1459
Dy 1309 ATGGCACCTGTGTAATCAAGCTCCAGGGGTACAGAGTCTCCGTACCAAGCAAAATAGTA 1368

Qy 1460 TATGGCTCAACATGATAGCCCCCAGATGGGAGTCTCTGCTCTGCCCCCAACAGCAG 1519
Dy 1369 TATGGCTTAAACATGAACGCCCCAGCCCGCATGGCAGTCCGGGCATGAATGCGCAACCAACC 1428


```
W58927 W97878"
<12013..12427
/note="match: GSS AQ035209 clone 2329M4"
13581..13604
/note="12 copies 2 mer aa 100% conserved"
complement(join(14705..17086,17420..17561,17912..18086,
19216..19510,21395..21487,23126..23419,28108..28279,
30441..30567,30670..30870,31290..31484,32709..32844,
33730..34601,34779..35170,36297..36474,36856..36996,
42469..42570,42743..42931,43316..43490,45011..45111,
46409..46581,48162..48265))
/genes="dJ1049G16.2"
/notes="match: CDNAS AF016031 AF012108 AF036892 AF010227
AF000581 AF044080 X97674 U39060 U80737 U64828 U64606
U56920 U59302 AJ000881 AJ000882 M60725 AJ005690; match:
ESTs AQ065268 AA488485 AA114092 AA065270 W84822 R28559
AA065272 AJ300819 T77368 AA065271 W84775 AA045419 AA150333
R21765 R66357 AA065269 W46433 AA384019 AA764263 AA811653
AA922415 AA828040 AA905298 AA729022 AA360136 R21718 W46488
H52110 AA678029 AA045313 H51992 AA767048 R67499 W20355
C05232 AA665470 AT146974 AA228716 AA230072 AA305762
AA494062 AA346492 AA346493 AA809279 AA738120 A1270423
AA488334 AA678419 AA883388 AA824532"
/evidence=not_experimental
complement(14705..48265)
/genes="dJ1049G16.2"
15605..15642
/note="19 copies 2 mer ac 95% conserved"
15732..15921
/genes="dJ1049G16.2"
/notes="match: STS G42545"
16761..17000
/genes="dJ1049G16.2"
/notes="match: STS G15161"
complement(join(17075..17086,17420..17561,17912..18086,
19216..19510,21395..21487,23126..23419,28108..28279,
30441..30567,30670..30870,31290..31484,32709..32844,
33730..34601,34779..35170,36297..36474,36856..36996,
42469..42570,42743..42931,43316..43490,45011..45111,
46409..46581,48162..48244))
/genes="dJ1049G16.2"
/notes="match: O09001"
/codon_start=1
/evidence=not_experimental
/product="dJ1049G16.2 (nuclear receptor coactivator)"
/protein_id="CAB40662.1"
/db_xref="PID:e1424047"
/db_xref="PID:g4584218"
/translation="MSGIGENLDPLASDRKRKRLCDTPGGLTSCGKRRRQESKY
IEELALISANLSDIDNFNVKDKAILKETVROIROIKQGGKTSINDDDDQKADVSS
TGQGVIDKSLGPLLQALQGLFVWNRDGNIVYSENVQYLOYKQEDLVNYSVNI
LHEDRKDFLKNLPKSTVNGSVTNETQRKSHFTFNCMLMKTPHDLIEDLVNINSPENR
QRYETMQCFALSPRAMEEGEDLQSCICVARRITGERTFPNPFSTFTRHDLQSK
VYNDTNSLSMRPGEEDIRRCIORFFSLNDQGSQSRHYQEVSDGIFSFTAYL
NGHATVPYRESLADGIVTAOTKSLFRNPVTNDRHGFVSTHFLQEQNGYRNPNP
VGOCIRPDMAGNSVSGMSKPNQGLQMPSSRAYGLADPSTTGOMSGRAGSSNIA
SLTGPQMGQSPSSQNNYGLNMSPPHGPSPLAPNQNMISPRNRGSKIASHQFS
PVGHPGPMASGGTGNHSSSSLSALQALISEGVTSLISTSSPGPKLDSNPMNI
TPQSKVNGKQSPGLGYFCQNPVSESSMCQNSRDHLSKESSEYGEAGNORGLPE
SGHKHLQLLITKSDDRGHSLTNSPLDSKESSVSVTSPGSCSTSGGVSSTNS
MHGSLLOEKHRLHLKLONGNSPAEVAIKTAEATGKDTSSITSCGDGNNYKQELSPK
KKENALLRYLLDRDDPSDALSKELQPOVEGVNKMSCQTSSTIPSSQEKDPKITE
TSRSGSDLDNLDAILDGLTSSFYNNISISGNSHLGKQKQFGTNSLQSGKSSQVQ
SIRPPYNRAVLSFVSVGSSPVVKNIASFMLPKQPMGLGNPRMDSNGYSSQSD
WQPNKAGRPNNSMGRPGDYNTSLRPAALGSIPTPLRSLNISIFGARVLQO
QQQLMQRPGISIPMGGANPYGQAASNLQSLWPDGMLSEQVSHGQNRPLRNSLD
DLVGPNSNLEQSDERALLDQHTLLSNTDATLEEDRALGIPELYNQCALPKPD
AFQGEAAVMQKAGLYGQTYPAQPGPMOGGHGLOGSPSFSNMNMNQGNFLQ
GMHPRANIMRPTWPKLRMLQLRLOGQFLNQSRQALELKNENTAGAAVMRPM
MQPQGGFLNAQMRASRLSHHFROORVAMMMQQQQQQQQQQQQQQQQQQQQQQ
QQQQQAFSPPPNTSPSNDHLLAGTTPMAPPPQPPQPPQPPQPPQPPQPPQPP
SPPNMMSSRMGSPQNPMQHPQAASIQSSEMKGWPSSGNLARNSSFSQQQFAHQGNP
```

```
AVYSVMHNGSSGHGMNMPMSPMGPMPDQKYC"
17102..17123
/note="11 copies 2 mer aa 100% conserved"
18527..18827
/genes="Alusg repeat: matches 1..297 of consensus"
19334..19631
/genes="dJ1049G16.2"
/notes="match: STS G09716"
19335..19418
/notes="4 copies 21 mer 93% conserved"
19335..19421
/notes="29 copies 3 mer tgc 93% conserved"
19621..19909
/notes="AlusX repeat: matches 1..293 of consensus"
19942..20232
/notes="Alusg repeat: matches 1..300 of consensus"
20337..20329
/genes="dJ1049G16.2"
/notes="match: 294499 chromosome 20 HindIII fragment"
complement(20260..20572)
/genes="dJ1049G16.2"
/notes="match: GSS AQ065701"
21387..21366
/notes="L2 repeat: matches 2632..2709 of consensus"
21744..22015
/notes="AluJb repeat: matches 17..308 of consensus"
22288..22586
/notes="AlusX repeat: matches 1..309 of consensus"
23503..23578
/notes="L2 repeat: matches 2679..2750 of consensus"

Query Match      24.1%   Score 1084.2; DB 10; Length 86453;
Best Local Similarity 86.5%; Pred. No. 5.1e-249;
Matches 1285; Conservative 0; Mismatches 23; Indels 177; Gaps 1;

Qy 1187 TTCCTTCAGAGAACAGATGATAGACAAACCAATCCTGTGGACAAGGGATT 1246
    |||
Db 35181 TTGTTTTCAGAGAACAGATGATAGACAAACCAATCCTGTGGACAAGGGATT 35122

Qy 1247 AGACCACTATGGCTGGATGCAACAGTTCGGTAGCGGCATGATGTCCGCCAACCA 1306
    |||
Db 35121 AGACCACTATGGCTGGATGCAACAGTTCGGTAGCGGCATGATGTCCGCCAACCA 35062

Qy 1307 GGCCTACAGATGCCAGCAGCAGGCGCTATGGCTTGCAGACCCCTAGCACCACAGGGCAG 1366
    |||
Db 35061 GGCCTACAGATGCCAGCAGCAGGCGCTATGGCTTGCAGACCCCTAGCACCACAGGGCAG 35002

Qy 1367 ATGAGTGGAGCTAGGTATGGGGGTTCACAGTAACATAGCTTCATTGACCCCTGGGCCAGGC 1426
    |||
Db 35001 ATGAGTGGAGCTAGGTATGGGGGTTCACAGTAACATAGCTTCATTGACCCCTGGGCCAGGC 34942

Qy 1427 ATGCAATCACCATTCTCTACCAGAACAACTATGGCTCAACATGAGTAGGCCCCCA 1486
    |||
Db 34941 ATGCAATCACCATTCTCTACCAGAACAACTATGGCTCAACATGAGTAGGCCCCCA 34882

Qy 1487 CATGGAGTCTCTGCTTGCCTCCCAACACAGCAGAAATCATGATTTCTCCTCGTAATCGT 1546
    |||
Db 34881 CATGGAGTCTCTGCTTGCCTCCCAACACAGCAGAAATCATGATTTCTCCTCGTAATCGT 34822

Qy 1547 GGGAGTCCAAAGATAGGCTCACATCAGTTTCTCCTCTGTTG----- 1587
    |||
Db 34821 GGGAGTCCAAAGATAGGCTCACATCAGTTTCTCCTCTGTTGTTGTTGACATT 34762

Qy 1587 ----- 1587
    |||
Db 34761 TCCTTTATTTTCTTTTAAAGTAATATTCTTCCATACTACTATAATAATCTCTGTAA 34702

Qy 1587 ----- 1587
    |||
Db 34701 AGTTAATCTATTTTATAATAGAAACTTTGAAGAAATTAAGTTGTATTTGTTACAGTGTCT 34642

Qy 1587 -----CAGGTGTGCACTCTCCCATGCCA 1609
    |||
```


[illegible]

Query Match	12.4%	Score 559.4	DB 10	Length 6156
Best Local Similarity	51.0%	Pred. No. 1.5e-123		
Matches 2197	Conservative	0	Mismatches 1836	Indels 276
Gaps				
64	CAGTTGCTGATGATATTC	AAGATGAGTGGATAGAGAAA	--CTTGGATCCACTGC	120
141	CAGTTGCTGATATGTT	CAAGATGAGTGGGATAGGAAATACCT	CTGACCCCTCCAG	200
121	CAGTGATTACGAAACG	CAAAATGGCATGTATCTCAGGACAAAGTCTT	ACCTGCGAG	180
201	GGCAGAGACAAGAAAG	CGCAAGAAATGCTCGACCAACTTGGACCCAGC	CCCCAAAAGNA	260
181	TGGTGAACGAGAGAG	CGGAGCAGGAAGATAATATGTAAGAAATGGCTGAGCTGAT	240	
261	CACTCAAAAAGTAACT	CGTGAACAGGAAAAATAATATATAGAAGAACTTGCAGAGTTGAT	320	
241	ATCTGCCAATCTTAGT	GCATATTCACAAATTCATGTCAACACAGATAAATGCGCATTTT	300	
321	TTTGTCAAAATTTAAT	GATATAGCAAACTTTAACTTCAAACTTGACAAATGTGCAATCTT	380	
301	AAAGAAACACGATAAG	CAGATACGTCACAAATAAAGACGAGAAA	---AACTATTTC	357
381	AAAGAAACGCTGAAG	CAAAATCGTCAGATCAAGAACAGAAAGACGACGCTGCCAA	440	
358	TGATGATGATGTTTCA	AAAAACCGATGATCTCTACAGGCGAGGAGTTATTGATAAAGA	417	
441	CATAGATGAAGTGCAG	AAGTCAGATGATCTCTACAGGCGAGGTTGTCATCGACAAGA	500	
418	CTCCTTAGGACCGCTT	TACTTCAGGCATTTGATGTTTCTTATTTGGTGAATCGAGA	477	
501	TGCGCTGGGGCCTATG	ATGATGTTGAGGCCCTTGATGGGTTCTTCTTTGTAGTAAACCTGGA	560	
478	GGCAACCATTTATTTG	TATCAGAAAATGCACAAATACCTGCAATATAAGCAAGAGA	537	
561	AGGCAACGTTGTTTGT	GCAGAGATGTCACAGTATCTAAGGTATAACCAAGAGA	620	
538	CTGTTTAACAAAGTGT	TACAAATATCTTACATGAAGAACAGAAAGATTTTCTTAA	597	
621	GCTGTGAACAAAAAGT	GTATAGCATTTGCATGTTGGGACACACGGAATTTGTCAA	680	
598	GAATTTAC---	CAAACTACAGTTAATGAGTTTCTGGACAAATAGCCCCCAAGACA	654	
681	AAACCTGTGCAAAAGT	CTATAGTAAATGGGGATCTTGGTCTGGCGAACCTCCGAGCG	740	
655	AAAAAGCATACATTAAT	TGATGCGGTATGTTGATGAACACCAACCATGATTTCTGGAAGA	714	
741	GAACAGCATACCTTCAT	TGTCGGATGTTGTTAAACCTTTACCTGATTCAGAGAGA	800	
715	CATAAGCCAGTCCCT	GAAATGGCCAGAGATATGAACAAATGCACTGCTTTGCCCTGTC	774	
801	GGGTATGATAACCGA	AGGCTCATCAGAAATATGAACATATGCACTGCTGCTGCTCTC	860	
775	TCAGCCAGAGCTATGAT	GAGNAGGGGAAGATTTGCAATCTTGTATCATCTGTGTGTC	834	
861	TCAACCAAGTCCATCA	AAAGAAAGAGAGAATTTGCACTGCTGCTGATTTGGGTGTC	920	
835	ACGCGCATTACTACAG	GAGAAAGCAATTTCCATCAAAACCTCGAGAGCTTTTATTACCA	894	
921	AAGAGAGTTCCTCAT	GAGNAGAACACCATCTTCCCTCATCAGAAAGTTTACTACTCG	980	
895	ACATGATCTTTTCAG	GAAGGTTGTCAATATAGATACAAATTCACAGATCTCCATGAG	954	
981	CCAGGATCTCCAAGG	CAAGATCATCGCTCTGATACCGACCATGAGAGCAGCATGAA	1040	
955	GCCTGGCTTTGAA	GATATATCCGAAGGTGATTCAGAGATTTTGTAGTCTTAATGATGG	1014	
1041	ACCAGGCTGGGAG	CACTGTTAAGAGGTTATTCAGAAAGTTCCATGCGCAGCATGAAG	1100	
1015	CGAGTCATGTTCCCA	---GAAAGCTACTATCAAGAGCTTATCTTAATGGCCATCGACA	1071	
1101	AGAACTGTGTCTTAT	GCTAAGAGGCATCATCATGAAGTACTGAGACAGAGGATGTGCATT	1160	
1072	AACCCCAAGTATPAT	CGATTTCTGTTGGCTGATGGAACCTATAGTGACCTGCACAGACAAAG	1131	

Db 2190 GTCTACATGAACCTCGCTCAAGGAGAAGCATAAAAATTTTGCACAGACTCTTGCAGGA 2249
Qy 2158 TGGGAATTCACAGCTGAGGTAGCAAGATTACTGCACAGCCACTGGGAAGACACACAG 2217
Db 2250 CAGAGTTCCTCCCTGGAGCTTGGCAAGTTAACAGAGAAGCCACAGGCAAAAGCTGAG 2309
Qy 2218 -----CAGTAAACTTCTTGTGGGACGGAATGTTGCAAGCAGGAGCAGCTAAG 2268
Db 2310 CCAGGAGTCCACGACGACAGCTCTCTGGATCAGAACTGACTATTAACRAGAGCCGGTGAG 2369
Qy 2269 TCCTAAGAAGAGGAGATAAATGCACTTCTTAGATACCTGTGGACAGGATGCTCTAG 2328
Db 2370 CCCCAGGAAGAAGAG---AATGCACTACTTCGCTATTGCTAGATAAAGATGACTATA 2426
Qy 2329 TGATGCACTCTTAAGAACTACAGCCCAAGTGAAGAGTGGACAATAAATGAGTCA 2388
Db 2427 AGATATGTTTACAGAAATACCCCAAACTTGAGAGACTGGACAGTAA---GACAGA 2483
Qy 2389 GTGCACAGCTCCAGCATTCCTAGCTCAAGTCAAGAGAAAGACCCCTAAATTAAGACAGA 2448
Db 2484 TCCGCGAGTACACAAATTAATAGCAATGAATGAACTGAGAGGAGGAGCTTGA 2543
Qy 2449 GACAAGTGAAGGAGTCTGGAGCTTGGATTAATCTAGATGCTATCTTGGTGTCTGAC 2508
Db 2544 GCCTGGTGACCAAGCTGGCAGTGAGCTGACAACTTGGAGGAGATTTGGATGATTGCA 2603
Qy 2509 TAGTTCGACTTTACATAAATTCATACC-----TCAA 2544
Db 2604 GAATAGTCAATTACACAGCTTTCCAGACAGAGGCGGCGCTGTGGATCAGT 2663
Qy 2545 TGGTAGTCACTGGGAGCTTAAGCAACAGGTGTTTCAAGGAATTAATCTCTGGGTTGAA 2604
Db 2664 TGACAAGCAAGCATCATCAATGACCTCATGCACTCACAGCTGAACAGCCCTGTAC 2723
Qy 2605 AAGTTCACAGTCTGTGAGTCTATTGCTCCTCCATATACCGAGAGTGTCTCTGGATG 2664
Db 2724 ACCTGTGGAGCCAGAAACAGCAGCTGGAATTTTCACAGAGCACTTTTAATAACCCACG 2783
Qy 2665 CCCTGTTTCTGTGCTCAAGTCCCTCAGTAAATAATACAGTGCCTTCCCATCTTACC 2724
Db 2784 ACCAGGCACTGGCAGGCTTATGCCAAACAGAGATTTACCAGTTGACATCACATGCA 2843
Qy 2725 AAAGCAACCATGTTGGTGGGAATCCAGAATGATGATAGTCAAGAAAT----- 2778
Db 2844 AAGCCCACTGTGTGGACCTTTCCCAACCATCAGAAACAGTAGTCCCTACTCAGTAT 2903
Qy 2778 -----ATGGCTCAAGTATGGGTAGGTTGGGCGCAACCGAATGTGACTGT 2817
Db 2904 ACCTCAGCCAGGAATGATGGGTAAATCAAGGGATGATAGGAAACCAAGGAATTTAGGGA 2963
Qy 2818 GACTCAGACTCTTCTCAGGAGACTGGGCTTACCAACTCAAAAGCGCGGAGAAATGGA 2877
Db 2964 CAGTAGCAGGAATGATGGTAACAGTCTCTCGGCTACTATGCCATCTGGGAATG 3023
Qy 2878 ACCTATGAATTCAAACTCCATGGGAAGACGAGGAGATTAATACTCTTTTACCAG 2937
Db 3024 GGCACCGCAGACTCGGCTGTGAGAGTCACTGTGCTGTACCACAGTGGCATGAACCG 3083
Qy 2938 ACCTGCTGGGTGGCTCTATTCCACATGTCCTCTTGGGTCTAATAGATACCAAGT--- 2996
Db 3084 GCCAGTCCAAGGAGTATGATTGGAACCCAGCAGCAGCATCCCATGAGGCCAGCAG 3143
Qy 2996 ---GCGAGACAGATTTGCAACAGCAGCAGATGCTTCAATGAGCCCTGGTGAAT 3051
Db 3144 CCAGCTTGCCCAAGACAGACGCTTCAGTCTCAGGTTCATGAATATAGGCGCATCTGAAT 3203
Qy 3052 CCCCATGGAATGGGGCTTAATCCCTATGGCAAGCAGCAGCATCTAAGCAACTGGGTTC 3111
Db 3204 AGAGATGAACATGGGGACCTCAGTATAGCCCAACAACAAGCTCTCCCAATCAGACTC 3263
Qy 3112 C---TGGCCCGATGGATTTGTCATGGAACAAGTTTCTCATGGCAGCTCAAAATAGGCC 3168
Db 3264 CCCATGGCTGAAGCATCTCTGCTATAGACCAGGCGTCTTTTGGCAGCAACACACAGCA 3323

Qy 3169 TCTTCTTAGGAATCCCTGGATGATCTTGTGGCCACCTTCCAACTTGAAGGCCAG 3228
Db 3324 GCCATTTGGCAGTTCTCCAGATGACTTCTATGTGCCATCTCGCAGCTGAGTCTCCAG 3383
Qy 3229 TGACAAAAGACATTAATGGACAGCTGCACACTCTTCTCAGCAACACAGATGCGACAGG 3288
Db 3384 TGATGAGGAGCTCTCTGGACAGCTGTATCTGGCTTGGGAATTTGAT-----GG 3437
Qy 3289 CCTGGAAGAAATGACAGAGCTTTGGGCAATTCCTGAACTTGTCAATCAGGGACAGCAT 3348
Db 3438 CCGTGGAGAGATTTGATAGAGCTTTAGGAATACCCGAATGCTGACGAGACCAAGCAT 3497
Qy 3349 AGAGCCCAACAGGATGCTTTCCAGGCCAAGACGACAGCTAATGATGGATCAAGAGC 3408
Db 3498 AGATCCAGAACAG-----TCTCAAGTCAGATTCCAACTATCATGCTGGAGCAAGAGC 3551
Qy 3409 AGGATTATATGACAGACATACCCAGCAGAGGGGCTTCCAATGCAAGGAGGCTTTTCATCT 3468
Db 3552 GCCCGTTTCCACAGCAGTATGCATCTCAGGCACAAATGCCAGGGTAGCTATTCTCC 3611
Qy 3469 TCAGGGAACATCACCATCTTTTAATCTATGATGAATCAGATGAACAGCAAGCAATTT 3528
Db 3612 CATGCAAGATCCAACTTTACACCATGGGACAGCGGCTAGTTATGCCACACTCCGTAT 3671
Qy 3529 TCCTCTCCAAGGAATGCCACCCAGCAGCAACATCATGAGACCCCGGACAAACACCCCAA 3588
Db 3672 CGAGCCAGA-----CCGGGCTCAGGCCACGGGCTAGTGCAGAACCCAGCCAAA 3722
Qy 3589 GCAACTTAAATGCACTTCAGCAGAGCTGCAGGGCCAGCAGATTTTGAATCAGAGCGG 3648
Db 3723 TCACTAAGACTTCACTTACGATCGCTTCCAAAGCAGCAGAGATCGCCAGC----- 3776
Qy 3649 ACAGCACTTGAATGAAATGGAACCCCTACTGCTGGTGGTGGTGGTGGTGGTGGTGG 3708
Db 3776 -----CACTTATGAATCAATCAGCAATGTTTCCAAATGTGAATTTGACTCTGAG 3824
Qy 3709 TATGATGAGCCCGCAGCAGGTTTCTTAATGCTCAAAATGGTCCGCAACGACGAGAGA 3768
Db 3825 GCCTGGAGTACCAACAGAGCACCTTATTANTGCACAGATGCTGGCCCGACAGACAGGGA 3884
Qy 3769 GCTGCTAAGTCACTACTTCCGACACAGAGGCTGCTATGATGATGACGAGCAGCAACA 3828
Db 3885 -----AATCCTGAACCAAGCATCT 3902
Qy 3829 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3888
Db 3903 TCGACAGAGACAAATGTCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3962
Qy 3889 CCAGGCTTTCAGCCCACTCTCTAATGTGACTGCTTCCCGCAGCAGTGGCTTTTGGC 3948
Db 3963 ACAAGGTTGAATATGACACCAAGCATGGTGGCTCTAGTGGTATGCCAGCACTATGAG 4022
Qy 3949 AGGACCCCAATGCCCAAGCTCTCCGCAACAGTTTCCATATCAACCAAAATTTATGGAAT 4008
Db 4023 CAACCTCGGATTCGCCAGGCAATGCACAGCTTCCATTTCTTCCAACTACGGAAT 4082
Qy 4009 GGGACAAACACAGATCCAGCCTTTGGTGGAGTGTCTAGTCTTCCCAATGCATGATGTC 4068
Db 4083 AAGTCAGCAAGCTGATCCAGGCTTTACTGGGCTACGACTCCCGCAGAGCCCACTTATGTC 4142
Qy 4059 GTCAGAATGGTCCCTCCAGAAATCCCATGATGACAAACCCGAGGCTGCATCCATCTA 4128
Db 4143 ACCCGAATGGCACAATACAGAGTCCCATGATGCAACAGTCTCAGGCCAACCCAGCCTA 4202
Qy 4129 TCAG---TCCTCAGAAATGAAGGGCTGGCCATCAGGAAATTTGGCCAGCAAGCTCCT 4185
Db 4203 TCAGGCCCTCCGACATAAATGGATGGCGCAGGGGAACATGGCGGAAACAGCATGTT 4262
Qy 4186 TTCCAGCAGCAGGTTTGGCCACCGAGGGAATCCTGCACTGTATAGTATG 4234
Db 4263 TTCCAGCAGTCCCCACACACTTTTGGCAGCAACACACACAGCATG 4311

Db 902 TCAGCCCAAGTCCATCAAGAGGCGAAGATTTGCAAGTCTGCTGATTTGTGTGGC 961
QY 835 ACSCCGCATTTACTACAGGAGAAAGACATTTCCATCAAGCCCTGAGAGCTTTATTTACAG 894
Db 962 ACAGAGAGTCCCATGAAGGAAAGACCACTCTCCCTCATCAAGAGCTTTTACCACCG 1021
QY 895 ACATGATCTTTTCAAGAAAGTGTCTCAATATAGATACAAATTTCACTGAGATCTCCCATGAG 954
Db 1022 CAGGACCTCAAGGACAGATCACTTCACTGGACATAGCACCATGAGAGCCGCCATGAA 1081
QY 955 GCCTGGCTTTGAAGATATATATCCGAAGGTGTATTACAGAGATTTTTTAGTCTAAATGATGG 1014
Db 1082 GCCGGGTGGGAAGATCTGTGTAAGAAGATGCATTCAGAAGTTCCACACACAGCATGAAG 1141
QY 1015 GCAGTC---ATGCTCCAGAAAGCTCACTATCAAGAGCTTATCTTAATGGCCATGCAGA 1071
Db 1142 GGAGTCTCTATCATATGCCAAGAGGATCAACCATGAAGTTCTGAGACAAGGTTGGCGTT 1201
1072 AACCCAGATATATCGATTCCTGCTGGTATGGAAGTATAGTACCTGACAGACAAAG 1131
Db 1202 CAGTCAGATCTATCGTTTTCTTTGCTGTGATGACATCTCGTTGCTGCACAAACCAAGAG 1261
QY 1132 CAACTCTTTCCGAAATCTGTAAACAAATGATCGACATGGCTTTGTCTCAACCCACTTCT 1191
Db 1262 CAAACTCATCCGTTCTCAGACTACTAATGAGCCTCAGCTTGTATATCTTTACACATGCT 1321
QY 1192 TCAGAGAAACAAANTGGATATAGACCAACCAACCTCTGTTGGACAGGGA----- 1245
Db 1322 TCACAGAGACGAANTGTATGTATGAATCCGGATCTGACTGGACAGCGATGGGAA 1381
QY 1245 -----TTAGACACACTATGCTGGATGCAACAGTTCCGTAGGCGGCATGATGTGCGC 1299
Db 1382 GCCATTGAATCCAAATAGTCTTAGAGCCCTGCCACAGGCCCTGTGCAGTGGGAACC 1441
QY 1300 AAACCAAGGCTTACAGATGCCGAGCAGAGGGCTATGCGTTGGCAGACCTTAGCACCAC 1359
Db 1442 AGGTCAGGACATGACCTCGGTAGCAATATAAATTTTCCCATGAATGGCCCAAGGAACA 1501
QY 1360 AGGCAGATGAGTGGAGCTAGGTATGGGGTTCCAGTAACATAGCTTCATTGACCCCTGG 1419
Db 1502 AATGGGCATCCCTATGGCAGGTTTGGTGTCTGGGGCA-----TGAACCATGT 1552
QY 1420 GCAGGATGCAATCACCATCTTCTTACCAGAACCAACTATGCGTCAACATGAGTAG 1479
Db 1553 GTCAGGATGCGAGCAACCACTCC---TCAGGATAGTAATGCACTCAAAATGAACAG 1609
1480 CCCCCACATGGGAGTCTGCTGCTTGGCCCAACACAGCAATATCATGATTTCTCCTCG 1539
Db 1610 TCCTCGCAAGCAGCCCGGCATGAACCGGGGCAAGCCAGCTCCGTGCTCTCCCCAAG 1669
QY 1540 TAATCG-----TGGAGTCAAGATAGCCCTACATCAGTTTCTCTCC 1581
Db 1670 GCAGCCATGAGCCCGGCGTGGCTGCGAGTCTCGATCCCAACCCAGTCAGTTTCCCC 1729
QY 1582 TGTTCAGAGTGTCACTCTCCCATGGATCTCTGCAATACCTGGAA---CCACAGCTT 1638
Db 1730 TGCAGNAGCTTGCATTCCTCTGGAGTTTGCACGACACAGGAATAGCCATAGTTA 1789
QY 1639 TTCACGAGCTCTCTCAGTCCCTGCAAGCCCATCAGTGAAGGTGTGGGACTTCCCTTTT 1698
Db 1790 TACCAACAGTTCCCTCAATGCACTGCAAGCCCTCAGGAGGGCATGGGGTCTCACTCG 1849
QY 1699 ATCTACTGTGATCACCAGGCCCAATTTGGATTACTCTCCC-----AATAT 1746
Db 1850 GTCTCTGCTGGCTTACCGGACCTAAATGGCAATTTGCAAAACTCCCCAGTTAATAT 1909
QY 1747 GAATATTACCAACCAAGTAAAGTAGCAATCAGGATTCACAGAGCTCTCTGGGCTTTTA 1806
Db 1910 GAATCTCCCCCTACTCAGCAAGATGGGAGCTTGGACTCCAAAGACTGTTTGGACTTTA 1969
QY 1807 TTGCGACCAAAATCCAGTGGAGGATTCATGTGTGCTAGTCAATAGC----- 1853

Db 1970 TGGGAGCCCTCAGAAGGTACAACCTGGACAAGCGCAGCTGCCATCTCTGAAGAACA 2029
QY 1853 -AGAGTACCTTCAGTCAGAAAGAAAGTAAGGAGGACAGTGTGAGGGGGCAGAGNATCA 1911
Db 2030 AAAGGGGCCAATGATTTCCAGCATGCCCGAGCGCCGAGGCGACAGGCTGAGGGACA 2089
QY 1912 AAGGGGTCTTTTGGAAAGCAAGGTCATAAAAATTTACTGCAAGTTACTTACCTGTTCTTC 1971
Db 2090 CAGCCGGCTGCTATGACAGCAAGGCGACCAAACTCTTGACAGTCTGCTGACCAAGTC 2149
QY 1972 TGATGACCCGGGTCTATCTCTTGACCAACTCCCGCCCTAGATTTCAAGTTGTAAGAANTC 2031
Db 2150 CGACCAAGATGGAGCTTACCCCTTGCCAGCTCTTGTGGACACAA----- 2197
QY 2032 TTCTGTAGTGTGACCAAGCCCTCTGGAGTCTCTGCTCTATCATCTGGAGGAGTATCTCTC 2091
Db 2197 -----ACAAGGACTCAACAGGAGCTTGCTGG 2224
QY 2092 TACATCCAATATGCTATGGGTCACTGTTACAAGAGAAAGACCGGATTTTGCACAAGTTGCT 2151
Db 2225 GCTGGGTCCAGCATGGCACCTCGCTCAAGGAGAAAGCATAGATTTTGCACAGACTCTT 2284
QY 2152 GCAGAAATGGAAATTCACACAGCTGAGGTAGCAAGATTTACTGCACAAGCACTGGGAAAGA 2211
Db 2285 ACAGGACAGCAGTTCCTGCTGGACTTGGCAAGCTGACAGCAGAAGCCACAGGCAAGA 2344
QY 2212 CACCAG-----CAGTATACTTCTTGTGGGACGAAATGTTGTCAAGCAGGAGCA 2262
Db 2345 GCTGAGCAGGAGTCCAGCAGCACAGCTCTGGGTGGAAGTGAAGTCTCAACACAGGAGCC 2404
QY 2263 GCTAAGTCTCAAGAAGAGGAATAATGCACTTCTTAGATACCTGCTGGACAGGATGA 2322
Db 2405 AGGAGCCCAAGAGAAAGAG---AATGCCACTACTGCGCTATTGCTCGACAAAGATGA 2461
QY 2323 TCTTAGTGTGACTCTCTTAAAGAACTACAGCCCAAGTGGAGGAGTGGCAATATAAT 2382
Db 2462 TACTAAGATATTGGTTTACCAGGAATAACCCCAAACTCGAGCGACTGGACAGTAAGAC 2521
QY 2383 GAGTCAGTCCACAGCTCCACCATTCTTCAAGTCAAGTCAAGGAGAAAGCCCTAAATTA 2442
Db 2522 AGATCTCGCAGTAACACAAAGTTAATGCTATGAAACTGTGAAGAG---GAGGTGAG 2578
QY 2443 GACAGAGCAAGTGAAGAGGAGTCTGGAGACTTGGATAATCTAGATGCTATTCTTGGTGA 2502
Db 2579 CTTTGAAGCCAGTACCCAGCTTGGCAGCAGCTGGACAACTTGGAGAGATTTTGGATGA 2638
QY 2503 TCTGACTAGTCTGACTTTTACAATAAATTCATATCTCTCAA----- 2544
Db 2639 TTTGCAAGACAGTCAAGTTACCACAGCTTTTCCAGACACAAAGCCAGGAGCTCCTACTGG 2698
QY 2544 -----ATGCTAGTCTATCTGGGACTAAGCAACAGGTTTTCAGGNACTAATTCCTGG 2598
Db 2699 GTCAGTTGCAAGCAAGCAGCATCATCAATGACCTCATGCAACTCAGAGTGCAGCAGTCC 2758
QY 2599 TTTGAAAGTCTCACAGTCTGTGAGTCTATTCGTCTCCATATAACCGAGCAGTGTCTCT 2658
Db 2759 CGTCCCACTGCGGAGCCCAAGAGCAGCACTGCGCATGTCACAGAGCACTTTTAATAA 2818
QY 2659 GGATAGCCCTGTTTCTGTTGGTCAAGTCTCCAGTAAAAAATATACAGTGTCTTCCCAT 2718
Db 2819 CCCACGACCGGCAACTGGCAGGTTATTGCCAAACCAAGAACTTACCACCTTGACATCAC 2878
QY 2719 GTTACCAAGCAACCCATGTTGGTGGGAATCCAAGAATGATGGTAGT-----CAGGAA 2773
Db 2879 TTTGCAAGCCCAACTGGTGGAGCTTTTCCCAACCAATCAGAAACAGTACGCCCTACTC 2938
QY 2774 AATTATGGCTCAAGTATGGTGGGCAACCCGAAATGTGACTGTGACTC----- 2823
Db 2939 AGTGATACCTCAGCCAGGAATGATGGTAACCAAGGATGCTAGGAGCCAGGAACCTT 2998
QY 2823 --AGACTCTCTCCTCAGGAGACTGGGGCTTACCAAACTCAAAAGCCGCA----- 2871
Db 2999 AGGGAACAATAGCACAGGAATGATTGGCAGCAGCACTTCCCGGCCCGCATGCTTCTG 3058

QY 2871 -GAATGGACCTATGAATTCACAACTCCATGGGAAGACACAGGAGGAGATTAATACTACTCT 2929
DB 3059 GGAATGGGACACACAGAGTCCAGCT--GTGAGAGTCACTTGTGCTGCTACCACTGGTGCC 3116
QY 2930 TTACCCAGACCTGCACTGGGTGGCTATT-----CCACATTGGCTCTTGGTCTAATA 2984
DB 3117 ATGACCCAGCAGTCAAGGAGGATGATTCGGAACCCACAGCAGATCCCATGGGA 3176
QY 2985 GCATACAGGTCGAGAC--AGTATTGCAACAGCAGCAGCAGATGCTTCAATGAGGCT 3043
DB 3177 GCCAACAGCAGCCTGGCCAAAGACAGATGCTTACAGTCTCAGTCTCATGAACATAGCCCT 3236
QY 3044 GGTGAATCCCCATGGGAATGGGGCTATATCCCTATGGCCAA---GCACAGCATCTAAC 3100
DB 3237 TCTGAGTTAGAGATGAACATGGGAGGACCTCAGTATATCAACAGCAGGCCCCCTCCGAC 3296
QY 3101 CAACCTGGGTTCTGCGCGATGGCATGTTGTCATGGGAACAGTTCATCATGCACTCAA 3160
DB 3297 CAACCTGCCCCGTGGCTGAGAGCATCTGCTATAGACGAGCATGCTTTCGCCAGCCAG 3356
QY 3161 AATAGGCTCTTCTTAGGAATTCCTGGATGATCTTGTGGGCCACCTTCCAACTGGAA 3220
DB 3357 AACAGCAGCCTTCGGCAGCTCCCTGATGACCTGCTGTGTCACATCCTCCAGCAGAG 3416
QY 3221 GCCCAGAGTGAAGAGCATATTGACACAGCTGACACATCTTCTCAGCAACACAGAT 3280
DB 3417 TCGCCAAAGCGATGAGGGGCTCTTCTTGACACAGCTGTATCTGCGCTTGCAGAACTTCGAT 3476
QY 3281 GGCAGAGCCTGGAAGAATTCACAGAGCTTGGCACTTCTGCACTGTCATCAGGGA 3340
DB 3477 -----GGCTTGGAGGATGATAGAGCTTGGGATACCAAGACTGTGTACGCCAGCAGC 3530
QY 3341 CAGGCATTAGAGCCCAACAGATGCTTCCAAAGGCCAAGACAGCAGCAGTATGATGAT 3400
DB 3531 CAAGCTGTGATGACAGCAG-----TCTCAAGTCAGGAGTCAGCATATGCTGAG 3584
QY 3401 CAGAGGAGGATTAATGACAGACATACCCAGCAGAGGGGCTTCCAAATGCAAGAGGC 3460
DB 3585 CAGAAAGCCCCGTTTCCACAGCAGTACGATCTCAGCAGCAAAATGGCCCCAGGCTGC 3644
QY 3461 TTTTCATCTCAGGACATCACTCTTTTACTCTATGATGAATCAGATGACACAGCAA 3520
DB 3645 TATAATCCCATG---CAGATCCAACTTTTACACATGGGACAGCGCCAAATACACC 3701
QY 3521 GGCATTTTCTCTCCAGGAATGACCCACAGCAGCAACATCATGAGACCCCGGACAAAC 3580
DB 3702 ACACCTCCGTATGACGCGCCGCGCCTCAGGCCACAGGATGTAC-----AGAAC 3755
QY 3581 ACCCCCAAGCACTAGATGAGCTTCAGCAGAGGCTGACGGCCAGCAGGTTTTTGAAT 3640
DB 3756 CAGCCAAACCAACTGAGACTTCAGCTTCAGCACCGCTCCAAAGCAGCAGAGATGCCAG 3815
QY 3641 CAGAGCCGACAGCCTGAATGAATGAATGGAACCCCTACTGCTGGTGGTCTCGGTG 3700
DB 3816 CCG-----CTTATGATCAGATCAGCAGTGTTCCTCAATGTGAACCTG 3857
QY 3701 ATGAGGCTATGATGACGCGCCAGCAGGCTTTCTTATGCTCAATGCTGCGCCCAAGCC 3760
DB 3858 ACTCTGAGGCTGGAGTGCCCACTCAGGCTCTATTAATGACAGATGCTGCGCCAGAGG 3917
QY 3761 AGCAGAGAGCTGCTAAGTATCATCTCCGACAAACAGAGGGTGGCTATGATGACGAGCAG 3820
DB 3918 CAGAGGGAATCCTCAACCAACCTTCG----- 3947
QY 3821 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3880
DB 3947 -----GCAGAGGAGAGATGACAGCAGAGGTCAGCAGCGGACTCTGATG 3989
QY 3881 CAGCAAAACCCAGGCTTCAGCCCACTCTCTAATGTGACTGCTTCCCCCAGCAGTGGTGG 3940
DB 3990 ATGAGAGGACAGGCTTGAATGTGACCCCAACAGCATGGTGGCTCCCGCTGCGCTACAGCA 4049

QY 3941 CTTTGGCAGGACCCACAAATGCCAACAGCTCTCCGCAACAGTTTCCATATATCAACCAAT 4000
DB 4050 GCCATGAGCAATCCCGGATCCCGAGCCAAATGCCAGCAGTTTCCATTTCTCCCGAAC 4109
QY 4001 TATGGAATGGGACACACACAGATCCAGCCTTTTGGTCCAGTGTCTAGTCTCCCAATGCA 4060
DB 4110 TACGGAATAAGTCAACAACCTGATCTGCTTTACTGGGGTACACTCCCGCAGAGTCT 4169
QY 4061 ATGATGTCGTCAAGAATGGTCCCTCCAGAAATCCCATGATGCAACACCCGAGCTGCA 4120
DB 4170 CTAATGTCCTCCCGATGSCATATCTCAGATGCCATGATGACAGCTCTCAAGCCAAC 4229
QY 4121 TCCATCTATCAG---TCTCAGAAATGAAGGCTGGCCATCAGGAAATTTGGCCAGAAC 4177
DB 4230 CCAGCTACCAAGCCACCTCAGCATGATGATGGACAGGGGAGCATGGGTGGAAAC 4289
QY 4178 AGCTCTTTTCCAGCAGCAGTTTCCACACAGGGGAATCCTGCGAGTGTATAGTAGT 4234
DB 4290 AGCATGTTCTCAGACAGTCCCGACACACTTTGGGCAACAGCAACACAGCATG 4346

RESULT 11

AF136943
LOCUS AF136943 4398 bp mRNA ROD 12-APR-1999
DEFINITION Rattus norvegicus transcriptional intermediary factor 2 (TIF2)
mRNA, complete cds.
ACCESSION AF136943
NID G4581052
VERSION AF136943.1 GI:4581052
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
REFERENCE 1 (bases 1 to 4398)
AUTHORS Leiris,J., Treuter,E. and Gustafsson,J.-A.
TITLE Mechanistic principles in NR box-dependent interaction between
nuclear hormone receptors and the coactivator TIF2
Mol. Cell. Biol. 18 (10), 6001-6013 (1998)
JOURNAL 98414616
MEDLINE 2 (bases 1 to 4398)
AUTHORS Leiris,J., Treuter,E. and Gustafsson,J.-A.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) Department of Biosciences, Karolinska
Institute, Haelsovaegen 7, Huddinge 14157, Sweden
FEATURES
Location/Qualifiers
1..4398 /organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..4398 /gene="TIF2"
1..4398 /db_xref="PID:G4581053"
/db_xref="GI:4581053"
/translation="MSGMGNTSDPSRAETRRKKECPDQLGSPKRSKTEKRNROENK
YIEELALIFANFDNIDNFKPDKCALIKETVKQIRIQEKQKAAANIDEVQKSDV
SSTGGVIDKDALGPMLEALDGFVFNLEGNVVFVSENVTVLRYNOELMKNKSVY
SILHVGDHTEFVKLLPKSMVNGSGTWGPPRRNSHTFNCRLMKVLPKLPDSEEGHDND
EAGQYETMQCFVAVSQPSIKEEGEDQSCCLICVARVPVKERALKPSPSEFTTRQDL
QGKTFIDTSMRDAMPWEDLVRCIKFHTQHEGESLSARKHHEVLRQGLAFS
QIYFSLSDGTLVAATQKSLKIRSQTNELPOLVISIHLHREQVNCVNMNPDLTQAMG
KPLSPSSSSSPARQAGKSGNPQVVALGSMNFMNPMNGPREQMSPMGFRFGSGMNHV
SGMAATTPQGSNVALKWNPSQSGNLPQPSVLPSPRHMSPCVAGSPRVPPSQFS
PAGSLHSPAGVCSSTGNSHSYTNSSNALALSEGHSVGLSPSLASPDLMKNSQNSP
VNMNPPLSKMSGLSKDCFLGIEPSEGTGQAQASCHPEEQKRPNDSSNPQASD
RAGHSRLHESKGTQKLLQLLTTKDQMEPSLPSSLSUDTNKSDTSLGPGSGTHGTS
LKEKHLHLRLQDSSVVDLAKLTAEATKELNDESGTAPGSEVTVKQBPAPKPKK
ENALLRYLLDKDDTKDGLPEITPKLERLDSKTDPSNKTLPASNTKLIANKTVKEEVSFSPDQ

Db 1849 ATGCCCCAGCGCCAGGAGGACAGGCTGAGGACACAGCGGCTGCATGAGACAA 1908
QY 1934 GGTCTATAAAAAATTACTGACAGTTACTTACCTGTTCTTCTGATACCGGGGTCTATCTCTCC 1993
Db 1909 GGGCAAAACCAACTCTCTACAGCTGCTCACCACCAAGTCCGACAGATGGAGCCTTCACCC 1968
QY 1994 TTGACCAACTCCCCCTTAGATTCAAGTTGTAAGATCTTCTGTTAGTGTCCACCCGCC 2053
Db 1969 TTGCCAGCTCTTTGTCGGACAA 1994
QY 2054 TCTGGAGTCTCCCTCTCATCTGAGGAGTATCTCTACATCCCAATATGTCATGGTCA 2113
Db 1994 -----ACAGGACTCCACAGCAGCTTCCAGGGCTGGTCCAGCATGGCACC 2043
QY 2114 CTGTTACAGAGAACACCGGATTTTGCACAAAGTTGCTGAGAAATGGGAATCACCAGCT 2173
Db 2044 TCGCTCAAGAGAGAGCAATAAGATTTTGCACAGACTCTTACAGACAGCAGCTCCCTGTG 2103
QY 2174 GAGTAGCCCAAGATTACTGACACAGCCACTGGCAAGACACCA-----GCAGTATA 2224
Db 2104 GACTTGGCCAAAGCTGACAGCAGAGGCACAGGCAAGAGCTGAATCAGGAGTCCAGTGGC 2163
QY 2225 ACTTCTTGTGGGACGGAATGTTGTCAAGCAGGAGCAGCTAAGTCTCTAAGAAAGAGGAG 2284
Db 2164 ACGGCTCTGGGTGAGAGTACTGTCAACAGAGGCCAGCGCCCAAGAAAGAG 2223
QY 2285 AATAATGCACTTCTTAGATACCTGTGGACAGGGATGATCTCTAGTATGACACTCTCTAA 2344
Db 2224 AAC-----GCACTACTGGCTATTTGTCGACAAAGATGATACTAAGATATTGTTACCG 2280
QY 2345 GAATACAGCCCCAAGTGGAGAGTGGCAATAAATGATGATGTCAGTCCACCCTCCACC 2404
Db 2281 GAATAACCCCCAACTCGAGCGGGTGGACAGTAA---GACAGATCCTGCCCAATACCA 2337
QY 2405 ATTCTTAGCTCAAGTCAAGAGAAAGACCTTAAATTAAGACAGACAGCAAGTGAAGAGGA 2464
Db 2338 AAGCTAATCCCATGAAACTGTGAAGGAGGAGTGAGCTTTGAGCCAGTGACAGCCT 2397
QY 2465 TCTGGAGACTTGGATTAATCTAGATGCTATTCTTGGTACTGACTAGTCTGACTTTTAC 2524
Db 2398 GGCAGCGAGCTGACAACTTGGAGAGATTTTGGATGATTTGAGATATGTCAGATTACCA 2457
QY 2525 AATAATCCATATCCCTCAATGTAGTATC-----TGGGGATAGCAACAGGTGTTTC 2579
Db 2458 CAGCTTTTCCAGACACAGCGCCAGGAGCTCTACTGGGTGAGTTGACRAGCAAGCCATC 2517
QY 2580 AAGGAATAATCTCTGGGTTTGAAGATTACAGTCTGTGCAAGTCTATTTCGTCCTCCAT 2639
Db 2518 ATCAATGACCTCATGCACTACGGCTGACAGCAGTCTGTCACTGTGGAGGCCAA 2577
QY 2640 ATAACCGAGCAGTGTCTCTGGATAGCCCTGTTT-----CTGTTGGCTCAAGTCCCTC 2690
Db 2578 AAAGCAGCACTGCGAATGTACAGAGCAGCTTTTAATAACCCAGCAGCAGGCAACTGGGC 2637
QY 2691 CAGTAAAAATATCAGTCTTTCCCATGTATACCAAGCAACCCATGTTGGGTGGGAATC 2750
Db 2638 AGGTTATTGCAAAACCAAGATTTTACCACCTTGACATCAGATTGCAAAAGCCCAACTGGTCT 2697
QY 2751 CAAGATGATGGATAGTACAGAAATATGGCTCAAGTATGGTGGGCCAAACCGAATG 2810
Db 2698 GGACCTTTCCCAACCATCAAGAAACAGTAGTCCATATCTAGTATACCTCAGCCAGGAA-- 2756
QY 2811 TGACTGTACTAGACTCTCTCTCAGGAGACTGGGCTTTACCAAACTCAAGGCCGGCA 2870
Db 2756 TGATGGGTACCAAGGATGTTAGGAGCCCAAGAACTTTAGGNAACATAGCACAGGAA 2815
QY 2871 GAATGGAACCTATGAATTTCAAACTCCA----- 2898
Db 2816 TGATGGCAGTAGCACTTCCCGGTCAGCATGCTTCTGGGGAATGGGCACACAGATC 2875
QY 2898 -----TGGGAAGACCCAGGAGAGATTAATACTCTTTTACCAGACCTGCACTGGGTG 2951
Db 2876 CAGCTGTGAGAGTCACTTGTGTGCTGCCACCACTGGTGCCATGAACCCGGCCCAATCCAAGGG 2935

QY 2952 GCTCTATT-----CCACATTGCCCTCTTTGGTCTTAATAGCATACAGGTGGGAGACCAG 3005
Db 2936 GCATGATTGGAACCCCAACGCAAGCATCCCATGGAGCAACAGCAGCCTGGCCAAA 2995
QY 3006 TATTGCAACAGCAGCAGCAGATGCTTCAATGAGGCTGGTNAATCCCATGGGAATGG 3065
Db 2996 GACAGATGCTGAGCCTCAGGTCATGAACATAGGGCTTCTGAGTTGGAGATGAACATGG 3055
QY 3066 GGCTAATCCCTATGGCCAA---GCAGCAGCATCTAACCACTGGGTTCTCTGCCCGCATG 3122
Db 3056 GAGGCCCTCAGTATATCAACAGCAGGCCCTCCCAACCAACTGCCCTGGCCCGAGA 3115
QY 3123 GCATGTTGTCATGGAACAAGTTTCTATGGCACTCAAAATAGGCTCTCTTTAGGAAT 3182
Db 3116 GCATCTGCTATAGACCAGGATCTTTGGCAGCCAGCAAGCATCCTTTGGCAGCT 3175
QY 3183 CCTGATGATCTTGTGGGCCACCTTCCAACTGGAAGCCAGAGTGAAGAAAGAGCAT 3242
Db 3176 CCCCCAGCAGCTGCTCTGTCCACATCTCTGACGCGAGTCCCAANGCAGGAGGGCGCTC 3235
QY 3243 TATTGACACAGCTGACACACTTCTCTCAGCAACACAGATGCGACAGGCTTGGAAATATG 3302
Db 3236 TTCTAGACAGCTCTATCTGGCTTTCGGAATTTGAT-----GSCCTTGAAGATTTG 3289
QY 3303 ACAGAGCTTTGGCATTTCTGAACTTGTCAATCAGGAGCAGGCAATTAGAGCCCAACAGG 3362
Db 3290 ATAGAGCTCTGGGATACCAAGAACTGTCAGCCAGAGCAAGCTGTGGATCCGAGCAG- 3349
QY 3363 ATGCTTTCCAAAGCCCAAGAGCAGCAGTAAATGATGATCAGAAAGCAGGATTAATGGAC 3422
Db 3349 -----TTCGAGTACAGAGTCCAGCATGATGCTGGAGCAGAAAGCCCTGTTTCCCAC 3403
QY 3423 AGACATACCCAGCAGCAGGCGCTCCCAATGCAAGAGGCTTTCATCTTACAGGCAATATC 3482
Db 3404 AGCAGTACGATCTCAGACACAAATGGCCCGGAGCTATAATCCATGCAAGATCCA- 3463
QY 3483 CATCTTTTAACTCTATGATGAATCAGATGAACCAAGCAAGCAATTTCTCTCCAGGAA 3542
Db 3463 -----AACTTCCACACCTGGACAGCGGCCAAATTTATCCACACTCCGAA 3508
QY 3543 TGCACCCAGCAGCCACATCATGAGACCCCGGACA-----AACACCCCAACAGCACTTA 3596
Db 3509 TCGACCCCGCGCGGCTCAGGCCACACAGGCTTGTGACRAGCAGCCCAACCACTGA 3568
QY 3597 GAATGAGCTTTCAGCAGAGCTGCGGGCCAGCAGTGTGATGATCAGAGCCGAGCAGC 3656
Db 3569 GACTTCAGCTCCAACTCGCTCCAAAGCAG-----CAGAACCCGCGCCGC 3616
QY 3657 TTGAATTTGAAAATGAAAAACCTTACTGCTGGTGTGCTGCGGTGATGAGGCTTATGTC 3716
Db 3617 T---AATGAACAGATCAGCGGCTTTTCCAAATGTGAATTTGACTTTGAGGCTTGGCGTGC 3673
QY 3717 AGCCCGCAGCAGGTTTCTTAATGCTCAATGCTGCCCAACCGCAGCAGAGAGCTGCTAA 3776
Db 3674 CAACT---CAGGCTCTATCAATGACAGATGCTGCCCGCAGAGGAGGAAATCCTTA 3730
QY 3777 GTCATCATTCCGCAACACAGAGGTGGCTATGATGATGACGAGCAGCAACAGCAGC 3836
Db 3731 -----ACCAGCATCTCCGGCAGA 3748
QY 3837 AGCAGCAGCAGCAGCAGCAACAGCAACAGCAGCAGCAACAGCAACCCAGGCT 3896
Db 3749 GACAGATGATCAGCAGCAGCAGGTCAGCAGGCACTTTGATGATGAGAGCAGAGGT 3808
QY 3897 TCAGCCCACTCTCAATGACTGCTTCCCGCAGCATGATGGCTTTTGGCAGGACCCA 3956
Db 3809 TGAATATGACCCCAAGCATGGTGGCTCCCACTGGTTTACCAGCAGCAGCATGAGCAACCCC 3868
QY 3957 CAATGCCAAGCTCTCCGCAACAGTTCCTCATATCAACCAATTTATGNAATGGGACAC 4016
Db 3869 GGATCCCGCAAGCCAAATGCCAGCAGTTCCTTCTCCGAACTACGGAATAGTCAGC 3928

```
QY 4017 AACCATCCAGCCCTTGGTCGAGTGTCTAGTCTCTCCCAATGCAATGATGTCGTCAGAA 4076
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3929 AACCTGATCTCGCTTACTGGAGCTACAACTCCCGAGAGCTCTTAATGTCCTCCCGGA 3988
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4077 TGGGTCCCTCCAGAAATCCCATGATGCAACACCCCGAGC---TGCATCCATCTATCATG 4133
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3989 TGGCACAATACGAGAGTCCCATGATGTCAGAGCTCTCAGGCCAAACCCAGCACCAGCCG 4048
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4134 CCTCAGAATGAAGGCTGGCCATCAGGAATTTGGCCAGGAACAGCTCTTTCCTCCAGC 4193
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4049 CTTACAGATTAACCGATGGGCACAGGAAGCATGGCGGAAACAGCATGTTCTCACAAC 4108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4194 AGCAGTTTCCCGACGAGGGAATCTCGAGTGTATAGTATG 4234
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4109 AGTCCCCACACACTTTGGGCAACAGCGAACACAGCATG 4149
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AF000582
US MUS musculus nuclear receptor coactivator protein 2 mRNA 25-JUN-1997
ORIGIN
INITIATION
ACCESSION AF000582
NID 92213816
VERSION AF000582.1 GI:2213816
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4771)
AUTHORS Torchia,J., Rose,D.W., Inostroza,J., Kamei,Y., Westin,S.,
Glass,C.K., and Rosenfeld,M.G.
TITLE The transcriptional co-activator p/CIP binds CBP and mediates
nuclear-receptor function
JOURNAL Nature 387 (6634), 677-684 (1997)
MEDLINE 97336097
REFERENCE 2 (bases 1 to 4771)
AUTHORS Torchia,J. and Rosenfeld,M.G.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1997) Medicine, Howard Hughes Medical Institute
at the University of California, San Diego, 9500 Gilman Drive,
C.M.M. 345, La Jolla, CA 92093-0648, USA
FEATURES
Location/Qualifiers
1. 4771
/organism="Mus musculus"
/db_xref="taxon:10090"
152..4543
/note="NCoA-2; nuclear receptor coactivator"
/product="nuclear receptor coactivator protein 2"
/protein_id="AAB61575.1"
/db_xref="PIR:G2213817"
/db_xref="GI:2213817"
/translation="MSGMGENTDPSRAETKRKCEPDQLGSPSKRSTERNRQENK
YIELAEIIDLKIVDFNFKPKCALIKETVQIKOEKAAANAEDEVKQSDV
SSTGGVIDKDALGPMLEALDGFVFNLEGSVFVFNVTQVLYRQELMKSVY
SILVAGDTEFVKLLPKSWNGSGWGEPPRRSHFNCMLVKPLPDLSEEGHDSQ
EAKQYEMQCFVSPKSIKEGEDLQSLIVVHEDPHEGKTNSSLRKLYHPGPP
RODFTGHYHRESREAGLGRSGKDAFRSSTHSMKGLYHMPRRHHVLRQGLAFS
QIYFSLSDGLVAOQTKSLIRQTTNEPOLVLSLHMLHREQNVCMNPDLTQOAMG
KPLNPISSSSPHALDCSGNPGQDMLSSNIFPMNGPKQOMGMPMRFVFGSGGMNV
SGMAATPOGSNYALKWNSPQSSPGMNGOASVLSLPRMSPGVAGSPRIPEPSQ
PAGNLHPVGVCSLDSKDCFLGIEPSKGTQGAESCHPKKQKQDNDSSMPQASGD
RAEGHSLRDLKQGLQLLTQDMEPLSPSLSDTNKDSGLPGSPHGTFS
LKEKHLIRLLQDSPPVLAKEATAEATKELSESSTAPGSEVTVKQEPAPFKK
ENALLRLDDEETLQNSLPQFDPTRGAPTGVKQAIINDLMOLTADSSPVPQA
GAKAALCMOSSFNPPGOLGRLLPYONLPDITLQSPGTAGFPPIRNSPYSVI
PDQSGKARFNPTASIPMGANSOLGQRMLOQSMNIGPELMMNGGPGYNOQAP
PNOTAPNPESILPDAQASFASNPQTFGSSPDDLLCPHPAESPDEGALLDLYAL
```

```

RNFEDGLEEIDRALGIPELYSQSQAVIDAEQFSOESSIMLEKQKPPVPOQYASQAQMAQ
GGYNPQDPNFHTMGORPNYTTLRMPRPGLRPTGIIVONOPNLRLOHRLQAOQNR
OPLMNOISSVSNVNLRLPGVPTQAPINAQMLAQRORILLNOHLRQOMQOQVQOETL
MQGQGLNVTPSVVAPAGLPAAMSNPRIPOANAQQPPFPNPGIISQDPGFGTATP
QSLMSPRMAHTQSPMWOQSQANPAYQPTSDMNGWAQSGMGSNMFSSQSPPHFGQQA
NTNMYNNMNIYSVMATNTGLSSMNQMTQNSMTSVTSVPTSLGSPMGPEQVNDPAL
RGNLFPNQLLGMIMIKQEGDASRKYC"
BASE COUNT 1292 a 1347 c 1228 g 904 t
ORIGIN
Query Match 10.5%; Score 474; DB 12; Length 4771;
Best Local Similarity 50.5%; Pred. No. 3.8e-103;
Matches 2180; Conservative 0; Mismatches 1845; Indels 289; Gaps 31;
QY 64 CAGTTGCTGATGATATCAAGATGAGTGGATTAGGAGAAAACHTTGGATCCACTGGCCAG 123
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 CAGTTGCTGATGATGTTCAAGATGAGTGGATGGGAGAAAACACCTCTGACCGCTCCAG 189
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 T---GATTTCAGGAAAACGCAAAATTCCTCATGTGATCTCCAGGCAAGGTCTTACCTGCAG 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 GCGAGAGACCAAAACGCAAGGAATGTCCGACACAGCTCGGACCCAGCCCCAAAGAG 249
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 TSGTGAATAACGAGACGGGAGCAGGAAAGTAAATATATTGAAGAATTGGCTGAGCTGAT 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 CACTGAGAAACGGAACCGCAGCAGGAGGAATAAGTACATAGAGGAGCTGGCCGAGCTGAT 309
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 ATCTGCCAATCTTAGTATATTGACAATTTCAATGTCAACCAGATAAATGGCGATTTT 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 310 CTTGCGAAACTTTAATGATATTGACAACTTCAACTTCAAACTTGACAAATGGCATCCT 369
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 AAAGAAACACGTAAAGACATACGTCAATATAAAGAGCAAGGAAA---AATATTTCAA 357
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 AAAAGAACTGTGAAGCAGATCCGCCAGATCAAGAGCAAGAAACGACGAGCTGCCA 429
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 358 TGATGATGATGTTTCAAAAGCCGATGATCTTCTACAGGGCAGGAGTATTGATAAGA 417
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 430 CATAGATGAAGTGCAGAAGTCAGATGTGTCGTCACGCGGCGAGGTGTCTATCGACAAG 489
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 418 CTCCTTAGGACCGTTTACTTTCAGGCATTTGGATGGTTTCTTATTTGGTGAATCGAGA 477
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 490 TGCCTAGGCGCCATGATGCTTGAAGCCCTCGATGGGTCTTCTTCTGTTGTAACCTGGA 549
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 478 GGCAACATCTTATTGTTGATCAGAAAATGTCAACAATACCTGCAATATAAGCAAGAGA 537
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 AGGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 609
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 538 CTGGTTAAACAAGTGTGTTTACAATATCTTACATGAAGAAGACAGAAAGGATTTTCTTAA 597
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 610 GCTGATGAACAAGAGTGTCTACAGCATCTCTGATGTGCGGGACACCACTGAATTTGTCAA 669
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 598 GAA---TTTACCAAAATCTACAGTTAATGAGTTTCTTGGACAATAAGCCCCAAGACA 654
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 670 GAACCTGCTGCGCAAGTCCATGTTGAATGGAGGATCTGTTGTTGTTGTTGTTGTTGTTGTT 729
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 655 AAAAGCCATACATTTAATTCCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 714
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 730 GAGCAGCTACCTTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 789
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 715 CATAAAGCCAGTCTCTGAAATGCCCCAGAGATATGAACAATGACAGTCTTGGCCCTGTC 774
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 790 AGGCATGATAGCCAGGAAGCCCATCAGAAATACGAGGCGATGAGTCTGCTGCTGCTGCTGCT 849
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 775 TCAGCCAGCAGCTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 834
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 850 TCAGCCCAAGTCCATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 908
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 835 ACGGCGCATTTACTACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 894
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 909 ACGAAG-ATCCCCATGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 967
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 895 ACATGATCTTTTCAGGAAAGGTTGTCATATATAGATACAAATTTACTGAGATCCTCCATGAG 954
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 968 CCAGGACCTCCAAAGCAAGATCACTTCTACTGGACACTACCACCATGAGAGCGGCCATGAA 1027
Qy 955 GCCTGGCTTTGAAGATATAATCCGAAGG-TGTATTCAGAGATTTTTTTAGTCTAAATGATG 1013
Db 1028 GCCGGCTGGGAAGATCTGGTAAGAAGATGCATTCAGAAGTTCCACACACAGCATGAAG 1087
Qy 1014 GGCAGTCATGCTCCAG---AAAGCTCAGTATCAAGAGCTTATCTTAATGGCCATGCA 1069
Db 1088 GGGAGTCTCTATCATATGCCAAGGAGGCATCACCATGAAGTTCGAGACAAGGTTGGCG 1147
Qy 1070 GAAACCCAGATATCGATTCTCGTTGGCTGATGGAATATAGTACGTGCACACACAA 1129
Db 1148 TTCAGTCAGATCTATCGTTTTTCTTCTGATGGCATCTCGTTGCTGCAACAACCAAG 1207
Qy 1130 AGCAAACTCTCCGAAATCTGTATCAAAATGATCGACATGGCTTTGTCTCAACCCATTC 1189
Db 1208 AGCAAACTCATCCGTTCTCAGACTACTAATGAGCCTCAGCTTGTAATATCTTTACACATG 1267
Qy 1190 CTTACAGAGAACAGAAATGGATATAGACCAACCAACCAATCTGTTGGACAAGGATTAAG 1249
Db 1268 CTTACAGAGAGCAAGATGTATGTATGAATCCGGATCTGACTGGACAAGCGATGGGG 1327
Qy 1250 CCACCTATG-----GCTGGATGCAACAGTTCTGGTAGCGGCATGAGTATGTCGCCAAGC 1303
Db 1328 AAGCCATGTAATCCATAGTCTTAGAGCCCTGCCCCACAGGCCCTGTGAGTGGGAAC 1387
Qy 1304 CAAGGCTTACAGATCCGAGCAGCAGGSCCTATGGCTTGGCAGAGCCCTAGCACACAGGG 1363
Db 1388 CCAGGTCAGGACATGACCCTCAGTAGCAATATAAATTTTCCCATGAATGGCCCAAGGAA 1447
Qy 1364 CAGATGAGTGAGTAGTATGGGGTTCCAGTAACTAGCTTCAATGACCCCTGGGGCCA 1423
Db 1448 CAAATGGGCATGCTATG---GGCAGGTTTGGTGGTCTGGGGGCATGAACCATGTGTCA 1504
Qy 1424 GGCATGAATCAGCATCTTCTACAGAACACAACTATGGCTCAACATGAGTAGCCCC 1483
Db 1505 GGCATGAAGCAACCACTCC---TCAGGGTAGTAACTATGCACTCAAAATGAACAGTCCC 1561
Qy 1484 CCACATGGGAGTCTGTGCTTCCGCCAAACCCAGCAGAAATATCATGATTTCTCCTCGTAAT 1543
Db 1562 TCGCAAGCAGCCCGGCATGAACCCGGGCAAGCCAGCTCCGTCTCTCCCAAGGCAG 1621
Qy 1544 CG-----TGGAGTCCAAAGATAGCCTCAATCAGTATTTCTCCTGTT 1585
Db 1622 GGCATGAGCCCGGCGTGGCTGGCAGTCTCGCATCCCAACCACTGAGTCTTCCCTGCA 1681
Qy 1586 CGAGGTGCACTCTCCCATGSCATCTTCTGCAATAGTGGAA---CCACAGCTTTCC 1642
Db 1682 GAAACTTGCATCTCCCTGTGGGAGTTTGCAGCAGCAGCAGGAAATAGCCATAGTATACC 1741
Qy 1643 AGCAGCTCTCAGTGCCTGCAAGCCATCAGTGAAGGTGTGGGACATTCCTCTTTATCT 1702
Db 1742 AACAGTTCCCTCAATGCATCGAAGCCCTCAGCGAGGCCATGGGTCTCACTCGGTCC 1801
Qy 1703 ACTGTGTCATCAGCAGGCCCAATTTGATTAACCTCC-----AATATGAAT 1750
Db 1802 TCGTGGCTTCCAGGACCTAAAAATGGCAATTTGCAAAACTCCCAAGTTAAATGAAT 1861
Qy 1751 ATTACCCAAACAGTAAGCAATAGCAATCAGATTCAGAGATCTCTGGGCTTTATTCG 1810
Db 1862 CTTCCCTCCACATCAGCAAGATGGGAAGCTTGGACTCCAAAGAGTGTTTTGGACTTTATGGG 1921
Qy 1811 GACCAAAAT-----CCAGTGGAGAGTTCAATGTGTCACTCAAAATAGCAGA 1855
Db 1922 GAGCCCTCAAAGGTACAACTGGCAAGCAGAGGCCACTGCCATCCGAAAAACAAAG 1981
Qy 1856 GATCACTCAGTACAAGAAAGTAAGGAGAGCAGTGTGTGGGGGGCAGAGAAATCAAGG 1915
Db 1982 GGGCCCAATGATTCAGCATGCCCCAGCGCCAGCGGGGACAGGGCTGAAGGACACAGC 2041
Qy 1916 GGTCTTTGGAAGCAAGAGTCAAAAAATTAAGTACTGAGTACTTACCTGTCTTCTGAT 1975
Db 2042 CGGTGTCATGACGCAAGGCGCAACCAACTCCTGCACTGCTGACCAACCAAGTCCGAC 2101

Qy 1976 GACCGGGTCTATCTCTCTTGACCAACTCCCCCTAGATTCAAGTTGTAAGAATCTTCT 2035
Db 2102 CAGATGGAGCCTTCCACCTTGCAGCTCTTGTGCGACACAAACAGGAC-----2153
Qy 2036 GTTAGTGTCAACGCCCTCTGGAGTCTCTCTCTTACATCTCGAGGAGTATCTCTACA 2095
Db 2153 -----TCGACAGGGAGCTTCCTCTGGGCT 2176
Qy 2096 TCCAATATGCATGGGTCACTGTTTAAAGAGAACCCCGGATTTTTCGCAAGTTGCTGAC 2155
Db 2177 GGGTCCAGCATGCACCTCGCTCAAGAGNAGCATTAAGATTTTGCACAGACTCTTACAG 2236
Qy 2156 AATGGGAATTCACAGCTGAGGTAGCCAAAGATTACTGCACAAGCCACTGGGAAAGACACC 2215
Db 2237 GACAGCAGTTCCCTGTGGACTTGGCCAAAGCTGACAGAGAAGCCACAGGCAAGAGCTG 2296
Qy 2216 AG-----CAGTATACTTCTTGTGGGACGGAATGTTGTCAAGCAGGAGCAGCTA 2266
Db 2297 AGCCAGGAGTCCAGCAGCAGACTCTCTGGTGGGAAGTGAAGTGTCTCAACAGGAGCCAGCG 2356
Qy 2267 AGTCTTAAGAAGAGGAGAAATAATGCACCTTCTTAGATACCTGCTGGACAGGATGATCCT 2326
Db 2357 AGCCCAAGAAAGAG---AATGCACACTCTGCGCTATTGCTCGACAAAGATGATACT 2413
Qy 2327 AGTATGCACCTCTCTAAAGAACTACAGCCCAAGTGGAAAGGAGTGGCAATATAAATGAT 2386
Db 2414 AAGATATTGGTTTACCGGAATAACCCCAACTCGAGCGACTGGACAGTAAGACAGAT 2473
Qy 2387 CAGTGACAGCTCCACCATCTCTAGTCAAGTCAAGTCAAGAGAAGACCCCTAAATTAAGACA 2446
Db 2474 CTTCCAGTAAACAAAGTTAAATGCTATGAAACTGTGAAGG---GGAGGTGAGCTTT 2530
Qy 2447 GAGCAAGTGAAGAGGATCTGGAGACTTGGATAATCTAGATGCTATCTTGGTGAATCTG 2506
Db 2531 GAGCCAGTGACCAAGCTGGCAGCGAGCTGGACAACCTTGGAAAGAGATTTTGGATGATTG 2590
Qy 2507 -----ACTAGTTCTGACTTTTACAATAATTCATATCTCTCAATAGTA 2549
Db 2591 CAGAACAGTCAGTTTACCACAGCTTTTCCAGACACAGGCCAGGAGCTCTCTACTGGGTCA 2650
Qy 2550 GTCATCTGGGACTTAAGCAACAGGTGTTTCAAGGAACATA-----2590
Db 2651 GTTGACAAGCAAGCATCATCAATGAGCTCATGCAACTCAGAGCTGACAGCAGTCCCGTC 2710
Qy 2590 -----TTCCTCTGGGTTGAAAAGTTCAAGTCTCTGAGTCTATTCGTCCTCCATATAAC 2644
Db 2711 CCACCTCGCGAGCCCAAGAGGAGCAGCTGTGATGTCAACAGAGCAGTTTTTAATAACCCA 2770
Qy 2645 CGAGCAGTGTCTCTGGATAGCCCTGTTTCT-----GTTGGCTCAAGTCTCCAGTAA 2696
Db 2771 CGACAGGCCAATCGGCAGGTATTGCCATACAGAACTTACCACCTTGACATCACATTG 2830
Qy 2697 AAAATATCAGTGTCTTCCCATGTTTACAAAGCAACCCATGTTGGGTGGGAATCCAAGAA 2756
Db 2831 CAAAGCCCAACTGTGTGTGGACCTTCCCACTCAGAAACAGTAGCCCTACTCAGTG 2890
Qy 2757 TGATGGATAGTCAGAAAATATGCTCAAGTATGGGTGGGCCAAACCGAATGTGACTG 2816
Db 2891 ATACCTCAGCAGGAATGATGGGTAAACAGGGATGTAGGAAGCCCAAGGAATTAGG 2950
Qy 2817 TGACTCAGACTCTCTCTCAGGAGACTGGGCTTTACAAACTCAAGGCCCGC---AGAA 2873
Db 2951 AACAAATGACAGAAATGATTGGCAGCAGCACTTCCCGGCCAGCATGCTCTCTGGGAA 3010
Qy 2874 TGAACCTATGAATTCAAACTTCCATGGGAAGACCCAGGAGGAGATTATAATCTTCTTAC 2933
Db 3011 TGGGCACACAGAGTACCAGCTGTGAGAGTACATTTGTGTGTCTACCACCTGGTGCCAAGA 3070
Qy 2934 CCACACCTGCACTGGGTGGCTCTATT-----CCCAATTCCTCTTCGGTCTTAATAGCA 2987
Db 3071 ACCGACAGTCCAAAGGAGGCAAGATTCCGGNAACCCCAAGCAGCATCCCCATGGGACC 3130

Query Match	7.38;	Score 328.4;	DB 10;	Length 4547;
Best Local Similarity	59.7%;	Pred. No. 2.4e-68;		
Matches 633;	Conservative 0;	Mismatches 406;	Indels 21;	Gaps 4;
QY	156	CTCCAGGACAAAGGTC	TTACCTGCAGTGGTGTGAAACGGAGCGGAGCAGGACGAGAAAGTAAT	215
DB	107	CGCCATGTGCACACT	TGGCATCAAGCAGGAAAGAGCGGAGGAGCAAGAAATAAT	166
QY	216	ATATTGAAGAAAT	TGGCTGAGCTGATATCTGCCAATCTTAGTGATATTGACAAATTCATG	275
DB	167	ATTTAGRAGAACT	AGCTAGCTGCTGCCAATCTTAGTGATTTGACATTTGACAGCTTGAGTG	226
QY	276	TCAAACGAGATAAT	TGTCGATTTTAAAGGAAACAGTAAGACAGATACGTCAAAATAAA--	334
DB	227	TAAACCCAGACAAAT	TGCAAGATTTTGAAGAAACAGTCGATCAGATACAGCTAATGAAGA	286
QY	334	-AGACRAGGAAACAT	TTTCCAAATGATGATGTTCAAAAGCCGATGTATCTCTTA	392
DB	287	GAATGGAAACAAGAAAT	CAACAATGATGACGATGTACAGAAATCAGACATCTCATCAA	346
QY	393	CAGGCGAGGAGT	TATTGATAAAGACTCCTTAGGACCGCTTTTACTTCAGGCATTTGGATG	452
DB	347	GTAGTCRAGGAGT	GTAGAAAAGAACTCTTTGGGACCTCTCTTTTGGAGCTTTGGATG	406
QY	453	GTTTCCTATTTTGGT	GAATCGAGGCAACATTTGATTTGTATCAGAAAATGTCACAC	512
DB	407	GATTTTCTTTTGT	TGAACCTGTGAAGGGAGAAATGTTATTTGTGCAGAGAAATGTAACA	466
QY	513	AATACCTGCAATATA	AGCAAGAGGACCTGGTTTACACAAGTGTTTACAAATCTCTTACATG	572
DB	467	GCTACTTAGTGTACA	ATCAGGAGGAATTAATGAATACGAGCGCTCTACAGCATACTGCACG	526
QY	573	AAGAAGCAGAAAGAT	TTCTTAAAGAT---TTACCAAAATCTACAGTTAATGGAGTTT	629
DB	527	TGGGGATCATGCAGA	ATTTGTGAAGAACTGTACCAAAATCACTAGTAATGGAGTTC	586
QY	630	CCTGGCAAAATGAGC	CCCCAAAGACAAAAGGCCATACATTTAATTCGGTATGTTGATGA	689
DB	587	CTTGGCCTCAAGAG	CAACACAGCAATACCCATACCTTTACTGCAGGATGCTAATTC	646
QY	690	AAACACACATGATAT	CTGTGGAAGACATAAAGCCAGTCTGAAATGGCCAGCAGATATG	749
DB	647	ACCTCCAGATGAGC	AGGG-----ACCGAGAACCAAGAGCTTGCCAGCGTTATG	697
QY	750	AAACAATGCAGTGT	TGCCCCTGTCTCAGCCACAGCTATGATGGAGAAAGGGAAGATT	809
DB	698	AAGTAATGCAGTGT	TCTACTGTGCAGCCAAATCAATTTCAAGAGGATGGAGAAAGATT	757
QY	810	TGCAATCTTGATGAT	CTGTGGCAGCCGCCATTACTACAGGAGAAAGACATTTCCAT	869
DB	758	TCCAGTCATGTCAT	TGTGTTATTCAGCGGCGATTACCTCGGCCCTC-----CAGCTATTA	811
QY	870	CAAAACCTGAGAGCT	TATTACCAGACATGATCTTTACGAAAGGTTGTCTCAATATAGATA	929
DB	812	CGGGGTAGAATCT	TATTATGACCAAGCAAGATACTACAGGTAATAATCATCTCTATTGATA	871
QY	930	CAAAATTCAGATCCT	CCATGAGGCCCTGGCTTTGAAGATTAATCCGAGGTTGATTTC	989
DB	872	CTAGTCCCTCAGAGCT	GTCTGGGAGAAATTTAGTGAGGAAGTGCATTT	931
QY	990	AGAGATTTTTTGTCT	TAAATCATGGCAGTCTATGTTCCACAAAAGCTCACATCAAGAAG	1049
DB	932	ATGCTTTTTTCCAC	CTCAGGCGAGACCAATCTTATGCCAGACAGCTGTTCCAAGAG	991
QY	1050	CTTATCTTAATGGC	ATCCAGAAACCCAGTATATCGATTTCTGTTGGCTGATGGAACATA	1109
DB	992	TGATGACTCGTGG	CACCTGCCCTCCAGCCCTCTATAGATTATTAATGATGGACAA	1051
QY	1110	TAGTACTGCACAG	ACAAAAGCAAACTCTTCCGAAATCTGTCAACAATGATCGACATG	1169
DB	1052	TGCTTAGCGGCCA	CCACCAAGTGTAAATTTCTACCTCTCAAGTCCGAGCATGCAACCTT	1111

```

Qy 1170 GCATTGTCTCACCCCACTTCCCTTCAGAGAGACAGATGG 1209
      | | | | | | | | | | | | | | | | | | | |
Db 1112 TCATCATGGGAATTCATATCATCGACAGGGACACAGTGG 1151

```

Search completed: September 18, 1999, 01:28:09
Job time: 22183 sec

M P S R E H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Search_pp protein - protein database search, using Smith-Waterman algorithm
on: Fri Sep 17 21:06:51 1999; MasPar time 56.90 Seconds
996.458 Million cell updates/sec
Tabular output not generated.

Title: >US-09-041-994-2
Description: (1-1415) from US09041994.pep
Sequence: 1 MSGLGENDPLASDSRRKL.....MNMNPMGMPGDPQKYC 1415

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 55.130; Variance 137.576; scale 0.401

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	9815	99.7	1417	2	receptor-associated c	0.00e+00
2	9778	99.3	1420	2	probable nuclear rece	0.00e+00
3	643	6.5	1061	2	steroid receptor coac	1.76e-76
4	232	2.4	848	2	aryl hydrocarbon rece	3.23e-14
5	239	2.4	853	2	aryl hydrocarbon rece	3.62e-15
6	222	2.3	313	2	alpha/beta-gliadin pr	7.10e-13
7	225	2.3	339	1	transcription initiat	2.82e-13
8	225	2.3	344	2	transcription factor	2.82e-13
9	230	2.3	357	2	omega secalin precurs	6.01e-14
10	229	2.3	357	2	omega secalin precurs	8.19e-14
11	230	2.3	644	2	transcription factor	6.01e-14
12	227	2.3	758	2	hypothetical protein	1.52e-13
13	231	2.3	805	2	aryl hydrocarbon rece	4.40e-14
14	226	2.3	905	1	regulatory protein SN	2.07e-13
15	231	2.3	4957	2	ALR protein - human	4.40e-14
16	231	2.3	5262	2	ALR protein - human	4.40e-14
17	221	2.2	609	2	odd-paired - fruit fl	9.65e-13
18	217	2.2	2175	2	homeotic protein cut	3.28e-12
19	208	2.1	139	2	Mopa box protein - mo	5.01e-11
20	210	2.1	360	2	Machado-Joseph disea	2.74e-11
21	205	2.1	467	1	involucrin - mouse	1.23e-10
22	202	2.1	539	2	hypothetical protein	3.02e-10
23	204	2.1	583	2	brain and muscle Ah r	1.66e-10

24	204	2.1	626	2	JC5405	brain and muscle Ah r	1.66e-10
25	205	2.1	648	1	JQ1150	protein kinase (EC 2.	1.23e-10
26	204	2.1	708	2	A53185	G-box-binding factor	1.66e-10
27	203	2.1	966	2	S23365	CYC8 protein - yeast	2.24e-10
28	204	2.1	1012	2	I53172	RAE-28 - mouse	1.66e-10
29	204	2.1	1081	2	S66736	transcription activat	1.66e-10
30	202	2.1	2038	2	A43742	female sterile homeot	3.02e-10
31	210	2.1	2414	2	A54277	transcription adaptor	2.74e-11
32	205	2.1	2440	2	S39162	CREB-binding protein	1.23e-10
33	205	2.1	2441	2	S39161	CREB-binding protein	1.23e-10
34	204	2.1	3144	2	A46068	Huntington disease-as	1.66e-10
35	199	2.0	181	2	JC5406	brain and muscle Ah r	7.37e-10
36	199	2.0	258	2	JC5407	brain and muscle Ah r	7.37e-10
37	196	2.0	271	2	PC4288	brain and muscle Ah r	1.79e-09
38	198	2.0	286	2	S07923	alpha/beta-gliadin pr	9.91e-10
39	200	2.0	326	2	D22364	alpha/beta-gliadin pr	5.47e-10
40	198	2.0	398	2	S21883	DNA-binding protein p	9.91e-10
41	200	2.0	613	2	S27770	hypothetical protein	5.47e-10
42	199	2.0	626	2	JE0270	Arnt-like PAS protein	7.37e-10
43	197	2.0	737	2	S47857	basic protein, cyto	1.33e-09
44	200	2.0	756	2	S60966	probable protein kina	5.47e-10
45	196	2.0	813	2	JC5809	HIF-1 alpha - rat	1.79e-09

ALIGNMENTS

RESULT 1
ENTRY T03443 #type complete
TITLE receptor-associated coactivator 3 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS T03443
REFERENCE Z14950
#authors Li, H.; Gomes, P.J.; Chen, J.D.
#journal Proc. Natl. Acad. Sci. U.S.A. (1997) 94:8479-8484
#title RAC3, a steroid/nuclear receptor-associated coactivator that is related to SRC-1 and TIF2.
#accession T03443
#status preliminary; translated from GS/EMBL/DBJ
#molecule_type mRNA
#residues 1-1417 ##label LIH
##cross-references EMBL:AF010227; NID:g2318005; PID:g2318006
##experimental_source cell line HeLa
GENETICS
#gene RAC3
#summary #length 1417 #molecular-weight 154535 #checksum 7159

Query Match	99.7%	Score	9815	DB 2	Length	1417			
Best Local Similarity	99.7%	Pred. No.	0.00e+00						
Matches	1413	Conservative	1	Mismatches	1	Indels	2	Gaps	1
Db	1	MSGLGENLPLASDSRKRKLPCDTPQGGLTCGCKRRRQESKYIEELAEIISANLSDID	60						
QY	1	MSGLGENLPLASDSRKRKLPCDTPQGGLTCGCKRRRQESKYIEELAEIISANLSDID	60						
Db	61	NFVVKPKCAILKETVRQIQIKQEGKTTISNDDVDKADVSSTGQGVKIDKSLGPLLQA	120						
QY	61	NFVVKPKCAILKETVRQIQIKQEGKTTISNDDVDKADVSSTGQGVKIDKSLGPLLQA	120						
Db	121	LDGFLFWNREANIVFVSENVTOYLQYKQEDLVNTSYNIIHEEDRKDFLKNLPKSTVNG	180						
QY	121	LDGFLFWNREANIVFVSENVTOYLQYKQEDLVNTSYNIIHEEDRKDFLKNLPKSTVNG	180						
Db	181	VSWTNETQKSHITFCNRMMLKTPHDILEDINASPENRQRYETMQCFALSQPRAMEEGE	240						
QY	181	VSWTNETQKSHITFCNRMMLKTPHDILEDINASPENRQRYETMQCFALSQPRAMEEGE	240						
Db	241	DLQSCMICVARRITTCGTFPPSPESFIRHDLGKGVNIDTNSLRSSMRPGFEDIIRRC	300						
QY	241	DLQSCMICVARRITTCGTFPPSPESFIRHDLGKGVNIDTNSLRSSMRPGFEDIIRRC	300						
Db	301	IQRFFSLNDQSQSKRHHYQAYLNGHAETPVYRFSIADGTIVTAQTKSLFRNPVTNDR	360						

```
QY 301 IORFFSLNDGQSWQKRHYQAYLNGHAETPVYRSLADGTTVTAQTKSLFRNPVTNDR 360
Db 361 HGFVSTHFLQREQNGYRPNPNVGGIIRPPMAGCNSSVGGMSMSPNOGLQMPSSRAYGLA 420
QY 361 HGFVSTHFLQREQNGYRPNPNVGGIIRPPMAGCNSSVGGMSMSPNOGLQMPSSRAYGLA 420
Db 421 DPSTTGQSGARYGSSNIASTLTPGQMGSPSSYQNNYGLNMSPPHPSGLAPNQNI 480
QY 421 DPSTTGQSGARYGSSNIASTLTPGQMGSPSSYQNNYGLNMSPPHPSGLAPNQNI 480
Db 481 MISPRNRSPKIAHQFSPVAGVHSPMASSGNTGNHSSSSLSALQAISEGVGTSLST 540
QY 481 MISPRNRSPKIAHQFSPVAGVHSPMASSGNTGNHSSSSLSALQAISEGVGTSLST 540
Db 541 LSSPGPKLDNSPNMNIQPSKVSQDQSKPLGFCYCDQNPVSSCMQSNRDLHSDKESKE 600
QY 541 LSSPGPKLDNSPNMNIQPSKVSQDQSKPLGFCYCDQNPVSSCMQSNRDLHSDKESKE 600
Db 601 SSVGAENQRPGLSKGHKKLLQLLTCSSDDRGHSSSLTNSPLDSSCKESSVSVTSPSGVS 660
QY 601 SSVGAENQRPGLSKGHKKLLQLLTCSSDDRGHSSSLTNSPLDSSCKESSVSVTSPSGVS 660
Db 661 STSGVSVSTSNHGSLLQEKHRIHLKLLQNGNSPAEVAKITAETATGKDTSSITSCGDGN 720
QY 661 STSGVSVSTSNHGSLLQEKHRIHLKLLQNGNSPAEVAKITAETATGKDTSSITSCGDGN 720
Db 721 VYKQQLSPKKENALLRYLLDRDDPSDALSKELQPOVEGVNDKMSOCTSTIPSSSQE 780
QY 721 VYKQQLSPKKENALLRYLLDRDDPSDALSKELQPOVEGVNDKMSOCTSTIPSSSQE 780
Db 781 KDPKIKTETSEEGSDLDNLDAILDGTSDFYNNSSISNGSHLTKQVFGTNSLGLK 840
QY 781 KDPKIKTETSEEGSDLDNLDAILDGTSDFYNNSSISNGSHLTKQVFGTNSLGLK 840
Db 841 SSQSVQSTRPNNRAVLDSPVSGSPVKNISAFPMPLPKPMLGGNPRMMDSOENYGS 900
QY 841 SSQSVQSTRPNNRAVLDSPVSGSPVKNISAFPMPLPKPMLGGNPRMMDSOENYGS 900
Db 901 SMGGNRRNVTVTQTPSSGDWGLPNKAGMEPWNNSMGRPGGDYNTSLPRALGGS IPT 960
QY 901 SMGGNRRNVTVTQTPSSGDWGLPNKAGMEPWNNSMGRPGGDYNTSLPRALGGS IPT 960
Db 961 LPLRSNSIPGARVPLVQQQQOQLMRPGETIPMGMGANPYCQAAASNQLGSWPDGMLSMQV 1020
QY 961 LPLRSNSIPGARVPLVQQQQOQLMRPGETIPMGMGANPYCQAAASNQLGSWPDGMLSMQV 1020
Db 1021 SHGTQNRPLLRNSLDLVGPPSNLGGSDERALLDQLHTLLSNTDQATGLEEIDRALGIPE 1080
QY 1021 SHGTQNRPLLRNSLDLVGPPSNLGGSDERALLDQLHTLLSNTDQATGLEEIDRALGIPE 1080
Db 1081 LYNQGOALEPKQDQAFQGEAAVYMDQKAGLYGQTYPAQGPMPGQGFHLGGQSPSFSNMN 1140
QY 1081 LYNQGOALEPKQDQAFQGEAAVYMDQKAGLYGQTYPAQGPMPGQGFHLGGQSPSFSNMN 1140
Db 1141 QNQQGNFPLQGMHPRANIMRPTNTPKOLRMQLQRLQGGQGLFQNSRQALELKMENPTA 1200
QY 1141 QNQQGNFPLQGMHPRANIMRPTNTPKOLRMQLQRLQGGQGLFQNSRQALELKMENPTA 1200
Db 1201 GGAAYVRPMQVQQGFLLNAQVQAQRSELLSHHFQQRVAMMMQOQQOQQOQQOQQOQQO 1260
QY 1201 GGAAYVRPMQVQQGFLLNAQVQAQRSELLSHHFQQRVAMMMQOQQOQQOQQOQQOQQO 1260
Db 1261 QQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 1320
QY 1261 QQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 1320
Db 1321 VSSPPNAMSSRMGFSQNPMMHQPAASTIQSSEMKGWPSGSLNARNSFSQOQFAHQGNP 1380
QY 1319 VSSPPNAMSSRMGFSQNPMMHQPAASTIQSSEMKGWPSGSLNARNSFSQOQFAHQGNP 1378
Db 1381 AVYSVHMNGSSGHGMQNMNPMGMPGPDQKYC 1417
```

```
QY 1379 AVYSVHMNGSSGHGMQNMNPMGMPGPDQKYC 1415

RESULT 2
ENTRY T03749 #type complete
TITLE probable nuclear receptor coactivator - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
ACCESSIONS T03749
REFERENCE Z15053
#authors Tanner, M.M.; Guan, X.Y.; Sauter, G.; Kallioniemi, O.P.;
#journal Science (1997) 277:965-968
#title AIB1, a steroid receptor coactivator amplified in breast and
#accession T03749
##status preliminary; translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-1420 #label ANZ
##cross-references EMBL:AF012108; NID:g2331249; PID:g2331250
GENETICS
#gene AIB1
#map_position 20
SUMMARY #length 1420 #molecular-weight 154891 #checksum 9406

Query Match 99.3%; Score 9778; DB 2; Length 1420;
Best Local Similarity 99.4%; Pred. No. 0.00e+00;
Matches 1411; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

Db 1 MSGLCENLDPLASDRKRLPCDTPGQGLTCSGKRRREQESKYEELAEILISANLSDID 60
QY 1 MSGLCENLDPLASDRKRLPCDTPGQGLTCSGKRRREQESKYEELAEILISANLSDID 60
Db 61 NFNVKPKCAILKETVROIQIRKEQKTIISNDQVQADSVSTGGQVTDKDSLGLPQLQA 120
QY 61 NFNVKPKCAILKETVROIQIRKEQKTIISNDQVQADSVSTGGQVTDKDSLGLPQLQA 120
Db 121 LDGFLFVYVNRGNIFFVSENVTQYLOKQEDLVNTSVYNIILHEEDRKDFLKNLPKSTYNG 180
QY 121 LDGFLFVYVNRGNIFFVSENVTQYLOKQEDLVNTSVYNIILHEEDRKDFLKNLPKSTYNG 180
Db 181 VSWNETORQKSHYFNCRLMKTKTPHDILEDINASPMPQRVETMOCFALSPRAMMEGE 240
QY 181 VSWNETORQKSHYFNCRLMKTKTPHDILEDINASPMPQRVETMOCFALSPRAMMEGE 240
Db 241 DLQSCMICVARRITTGERTFPPSPESFTRHDLGKVVYNIIDTNSLRSSMRPFGEDIIRRC 300
QY 241 DLQSCMICVARRITTGERTFPPSPESFTRHDLGKVVYNIIDTNSLRSSMRPFGEDIIRRC 300
Db 301 IORFFSLNDGQSWQKRHYQAYLNGHAETPVYRSLADGTTVTAQTKSLFRNPVTNDR 360
QY 301 IORFFSLNDGQSWQKRHYQAYLNGHAETPVYRSLADGTTVTAQTKSLFRNPVTNDR 360
Db 361 HGFVSTHFLQREQNGYRPNPNVGGIIRPPMAGCNSSVGGMSMSPNOGLQMPSSRAYGLA 420
QY 361 HGFVSTHFLQREQNGYRPNPNVGGIIRPPMAGCNSSVGGMSMSPNOGLQMPSSRAYGLA 420
Db 421 DPSTTGQSGARYGSSNIASTLTPGQMGSPSSYQNNYGLNMSPPHPSGLAPNQNI 480
QY 421 DPSTTGQSGARYGSSNIASTLTPGQMGSPSSYQNNYGLNMSPPHPSGLAPNQNI 480
Db 481 MISPRNRSPKIAHQFSPVAGVHSPMASSGNTGNHSSSSLSALQAISEGVGTSLST 540
QY 481 MISPRNRSPKIAHQFSPVAGVHSPMASSGNTGNHSSSSLSALQAISEGVGTSLST 540
Db 541 LSSPGPKLDNSPNMNIQPSKVSQDQSKPLGFCYCDQNPVSSCMQSNRDLHSDKESKE 600
QY 541 LSSPGPKLDNSPNMNIQPSKVSQDQSKPLGFCYCDQNPVSSCMQSNRDLHSDKESKE 600
Db 601 SSVGAENQRPGLSKGHKKLLQLLTCSSDDRGHSSSLTNSPLDSSCKESSVSVTSPSGVS 660
QY 601 SSVGAENQRPGLSKGHKKLLQLLTCSSDDRGHSSSLTNSPLDSSCKESSVSVTSPSGVS 660
```

```
QY 601 SSVGAENQRPGLSKGHKLLQLLTCCSDDRGHSLTNSPLDSSCKESSVSVTSPGCVS 660
Db 661 SSTGGVSSSTNMGSLLOEKHRLHLKLLQNGNSPAEVAKITABATKDKTSITSCDGN 720
QY 661 SSTGGVSSSTNMGSLLOEKHRLHLKLLQNGNSPAEVAKITABATKDKTSITSCDGN 720
Db 721 VVKEQLSPKKENALLRYLLDRDDPSDALSKELOQVEGVNKMSSOCTSSSTIPSSOE 780
QY 721 VVKEQLSPKKENALLRYLLDRDDPSDALSKELOQVEGVNKMSSOCTSSSTIPSSOE 780
Db 781 KDKPIKTETSEGGDLNDLDAIIGDLSSTDFYNNSSSSNGSHLGTQQVFGQFNSLGLK 840
QY 781 KDKPIKTETSEGGDLNDLDAIIGDLSSTDFYNNSSSSNGSHLGTQQVFGQFNSLGLK 840
Db 841 SSQSVQSIIRPYNRAVSLDSPVSGSSPPVKNIISAFPMPLPKQPLMGGNPRMDSQENYGS 900
QY 841 SSQSVQSIIRPYNRAVSLDSPVSGSSPPVKNIISAFPMPLPKQPLMGGNPRMDSQENYGS 900
QY 901 SMGGPNRNVTVTQTPSSGDWGLPNKAGRMPEMNSNMGPRGGDYNTSLRPALGGSIPT 960
QY 901 SMGGPNRNVTVTQTPSSGDWGLPNKAGRMPEMNSNMGPRGGDYNTSLRPALGGSIPT 960
Db 961 LPLRSNIPGARPYLQOQQOQMLQMRPGEIPMGANGPYGQAAASNLQSGWPDGMLSEQV 1020
QY 961 LPLRSNIPGARPYLQOQQOQMLQMRPGEIPMGANGPYGQAAASNLQSGWPDGMLSEQV 1020
Db 1021 SHGTQNRPLRNSLDDLVGPPSNLEGGSDERALLDQLHTLLSNTDATGLBIDRALGIPE 1080
QY 1021 SHGTQNRPLRNSLDDLVGPPSNLEGGSDERALLDQLHTLLSNTDATGLBIDRALGIPE 1080
Db 1081 LVNQGQALEPKQDAFOGQEAAYVMDQKAGLYGQTPYPAQGPMPQGGFHLQSGPSFNSMN 1140
QY 1081 LVNQGQALEPKQDAFOGQEAAYVMDQKAGLYGQTPYPAQGPMPQGGFHLQSGPSFNSMN 1140
Db 1141 QMNOQGNFPLQGMHPRANIRPTNTPKQLRMQLQOQFLNQSRQALELKNENPTA 1200
QY 1141 QMNOQGNFPLQGMHPRANIRPTNTPKQLRMQLQOQFLNQSRQALELKNENPTA 1200
Db 1201 GGAAMVRPMQPOQGFFLNAQVARSRELLSHHFRQORVAMMMQOQOQOQOQOQOQOQ 1260
QY 1201 GGAAMVRPMQPOQGFFLNAQVARSRELLSHHFRQORVAMMMQOQOQOQOQOQOQOQ 1260
Db 1261 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 1320
QY 1261 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 1315
Db 1321 FGRVSSPPNMMSSRMGSPQNMHQPAAASISYOSSEKMGWPSGNLARNSSFSQOQFAHQ 1380
QY 1316 FGRVSSPPNMMSSRMGSPQNMHQPAAASISYOSSEKMGWPSGNLARNSSFSQOQFAHQ 1375
Db 1381 GNPAYSVMHVHNGSSGHGMQNMNPMPSGMMPGPDQKYC 1420
QY 1376 GNPAYSVMHVHNGSSGHGMQNMNPMPSGMMPGPDQKYC 1415
RESULT 3
ENTRY #type complete
TITLE steroid receptor coactivator 1 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change
30-May-1997
ACCESSIONS A57620
REFERENCE A57620
#authors Onate, S.A.; Tsai, S.Y.; Tsai, M.J.; O'Malley, B.W.
#journal Science (1995) 270:1354-1357
#title Sequence and characterization of a coactivator for the steroid hormone receptor superfamily.
#cross-references MUID:96085063
#accession A57620
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-1061 #label ONA
KEYWORDS steroid hormone receptor
```

```
SUMMARY #length 1061 #molecular_weight 114166 #checksum 398
Query Match 6.5%; Score 643; DB 2; Length 1061;
Best Local Similarity 36.2%; Pred. No. 1.76e-76;
Matches 113; Conservative 89; Mismatches 89; Indels 21; Gaps 18;
Db 456 MDGANVSTVTKSEILPASLOSATARTSRNLN-RLPELELEAIDNQFGQPGTGOI-PWTN 513
QY 954 LGSISPTPLRNSIPGA-REVLIQQOQOQMLQMRPGEIPMGANGPYGQAAASNLQSGWPD 1012
Db 514 NTVTAIINQSR-SEDOCI-SGQDELICPPTTVEGRNDEKALLEGVFLSGKDELAEL 571
QY 1013 GMLSEQVSGTGNRPLRNSLDDLVGPPSNLEGGSDERALLDQLHTLLSNTDATGLEEI 1072
Db 572 DRALGIDKLV-QGGGLDVLSERPFPQOATPPLIMEERNLYSQYSPFPPTANLPSPFQ 630
QY 1073 DRALGIPELVNOGQALEPKQDAFOGQEA--VMDQKAGLYGQTPYPAQGP-P-MGGF 1127
Db 631 MYRQKPSLGTMPVQVTPPRGAFSPGCMQPRQTLNRRPA-APNQLRLQLOQRLOGQOOLI 689
QY 1128 LOGQSPSFNSMNMNQO-OGNF-PLQGMHPRANIRPTNTPKQLRMQLQOQRLOGQO-FL 1184
Db 690 HQNRQAI-LNOFAATAPVGINMSGMQOQITPOPPLNAQMLAORRELYSQOHRQ-R-QL 746
QY 1185 NQSRQALELKNENPTAGCAAVMRPMQPO---QGFLNAQVARSRELLSHHFRQORVAM 1241
Db 747 IQQORAMLMRQQ 758
QY 1242 MMQOQOQOQOQOQ 1253
RESULT 4
ENTRY #type complete
TITLE aryl hydrocarbon receptor - human
ALTERNATE_NAMES Ah receptor
ORGANISM #formal_name Homo sapiens #common_name man
DATE 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change
10-Sep-1997
ACCESSIONS S59514; JX0356; S41124
REFERENCE S59514
#authors Dolwick, K.M.; Schmidt, J.V.; Carver, L.A.; Swanson, H.I.; Bradfield, C.A.
#journal Mol. Pharmacol. (1993) 44:911-917
#title Cloning and expression of a human Ah receptor cDNA.
#cross-references MUID:94067047
#accession S59514
#status preliminary
#molecule_type mRNA
#residues 1-848 #label DOL
#cross-references EMBL:L19872; NID:g416141; PID:g416142
REFERENCE JX0356
#authors Ema, M.; Matsushita, N.; Sogawa, K.; Ariyama, T.; Inazawa, J.; Nemoto, T.; Ota, M.; Oshimura, M.; Fujii-Kuriyama, Y.
#journal J. Biochem. (1994) 116:845-851
#title Human arylhydrocarbon receptor: Functional expression and chromosomal assignment to 7p21.
#accession JX0356
#molecule_type mRNA
#residues 1-4,'G',6-356,'H',358-460,'N',462-631,633-848 #label EMA
#experimental_source lung
REFERENCE S41124
#authors Itoh, S.; Kamataki, T.
#journal Nucleic Acids Res. (1993) 21:3578
#title Human Ah receptor cDNA: analysis for highly conserved sequences.
#cross-references MUID:93347997
#accession S41124
#status preliminary
#molecule_type mRNA
#residues 1-404,'N',406-797,'E',799-806,'FK', #label ITO
COMMENT This protein is involved in the process in which 2,3,7,8-tetrachlorodibenzo-p-dioxin exerts its effects.
```

[illegible]


```
#title      Cloning of a transcriptionally active human TATA binding
#cross-references MUID:90302010
#accession   A34831
#status      not compared with conceptual translation
#molecule_type DNA
#residues    1-17,'N',19-186,'R',188-339 #label KAO
REFERENCE   S10944
#authors     Hoffmann, A.; Sinn, E.; Yamamoto, T.; Wang, J.; Roy, A.;
              Horikoshi, M.; Roeder, R.G.
#journal     Nature (1990) 346:387-390
#title       Highly conserved core domain and unique N terminus with
              presumptive regulatory motifs in a human TATA factor
              (TFIID)
#cross-references MUID:90326195
#accession   S10944
#molecule_type mRNA
#residues    1-91,96-339 #label HOF
#cross-references EMBL:X54993; NID:g37065; PID:g37066
GENETICS
#gene        GDB:TBP; GTF2D1
#map_position 6q27-6q27
#CLASSIFICATION #superfamily transcription initiation factor IID
KEYWORDS     alternative splicing; DNA binding; transcription initiation
FEATURE
55-95        #region glutamine-rich
SUMMARY      #length 339 #molecular-weight 37698 #checksum 4870
Query Match      2.3%; Score 225; DB 1; Length 339;
Best Local Similarity 38.5%; Pred. No. 2.82e-13;
Matches 42; Conservative 16; Mismatches 49; Indels 2; Gaps 2;
Db 18 OGAMTPGIPSPMPYGTG-LTPQIONTNLSILEEQRRQQQQQQQQQQQQQQQQQQQQ 76
QY 1196 ENPTAGGAANVRMPQPGFLNAQVARSR-ELLSHHFRRQVRVMMQMQQQQQQQQQ 1254
Db 77 QQQQQQQQQQQQQQQQQAARAAVQSTSQATQGTSGAQPQLFHSQ 125
QY 1255 QQQQQQQQQQQQQQQQAFSPPNVTASPSMDGLLAGTPTMPQAPQPPFYQ 1303
RESULT 8
ENTRY      I60128 #type complete
TITLE      transcription factor IID - human
ORGANISM   Homo sapiens #common_name man
DATE       04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change
ACCESSIONS I60128
REFERENCE   Kao, C.; Lieberman, P.; Schmidt, M.; Zhou, Q.; Pei, R.; Berk,
              A.J.
#journal    Science (1990) 248:1626
#title      Cloning of the human TATA binding factor: Expression of a
              transcriptionally active TFIID protein.
#accession  I60128
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues   1-344 #label RES
#cross-references GB:M34960; NID:g339493; PID:g339494
CLASSIFICATION #superfamily transcription initiation factor IID
FEATURE
55-109      #region glutamine-rich
SUMMARY      #length 344 #molecular-weight 38409 #checksum 7457
Query Match      2.3%; Score 225; DB 2; Length 344;
Best Local Similarity 38.5%; Pred. No. 2.82e-13;
Matches 42; Conservative 16; Mismatches 49; Indels 2; Gaps 2;
Db 18 OGAMTPGIPSPMPYGTG-LTPQIONTNLSILEEQRRQQQQQQQQQQQQQQQQQQQQ 76
QY 1196 ENPTAGGAANVRMPQPGFLNAQVARSR-ELLSHHFRRQVRVMMQMQQQQQQQQQ 1254
```

```
Db 77 QQQQQQQQQQQQQQQQQAARAAVQSTSQATQGTSGAQPQLFHSQ 125
QY 1255 QQQQQQQQQQQQQQQQAFSPPNVTASPSMDGLLAGTPTMPQAPQPPFYQ 1303
RESULT 9
ENTRY      S18235 #type complete
TITLE      omega secalin precursor (clone pSec1B) - rye
ORGANISM   Secale cereale #common_name rye
DATE       16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change
ACCESSIONS S18235; S19777
REFERENCE   Hull, G.A.; Halford, N.G.; Kreis, M.; Shewry, P.R.
              Plant Mol. Biol. (1991) 17:1111-1115
#journal    Isolation and characterisation of genes encoding rye
#title      prolamins containing a highly repetitive sequence motif.
#cross-references MUID:92032773
#accession  S18235
#molecule_type DNA
#residues   1-357 #label HUL
#cross-references EMBL:X60294; NID:g21201; PID:g21202
GENETICS
#gene        sec1
#CLASSIFICATION #superfamily gliadin
FEATURE
1-19         #domain signal sequence #status predicted #label SIG\
20-357       #product omega secalin #status predicted #label MAT
SUMMARY      #length 357 #molecular-weight 41541 #checksum 3278
Query Match      2.3%; Score 230; DB 2; Length 357;
Best Local Similarity 26.4%; Pred. No. 6.01e-14;
Matches 74; Conservative 55; Mismatches 133; Indels 18; Gaps 15;
Db 74 PTPIQPQPPQRPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 131
QY 1079 PELVNCQALEPK-QDAFQGEAAVMDQKAGLYGQT-YFAQGPMPGGFHLCQSPS-- 1134
Db 132 FPQPPQGIIPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 191
QY 1191 LELKMNPTAGGAANVRMPMPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 1248
QY 1135 FNSMMNQW-NQNGN--FPLQGMHP-RANIMRPTNTPKLRMLQQRLOGQQLNQSRQA 1190
Db 192 FPQPPQGIIPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 250
QY 1191 LELKMNPTAGGAANVRMPMPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 1248
Db 251 VPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 310
QY 1249 QQQQQQQQQ-QQQQQQQ-QQQQTQAFSPPNVTASPSMDGL--LAGPTMPQAPQPP 1304
Db 311 QPPFPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 349
QY 1305 NYGNGQPPDPAFGRVSSPPNMMSSRMGSPSONPMQHPQA 1344
RESULT 10
ENTRY      S18236 #type complete
TITLE      omega secalin precursor (clone pSec2B) - rye
ORGANISM   Secale cereale #common_name rye
DATE       16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change
ACCESSIONS S18236; S19778
REFERENCE   Hull, G.A.; Halford, N.G.; Kreis, M.; Shewry, P.R.
              Plant Mol. Biol. (1991) 17:1111-1115
#journal    Isolation and characterisation of genes encoding rye
#title      prolamins containing a highly repetitive sequence motif.
#cross-references MUID:92032773
#accession  S18236
#molecule_type DNA
#residues   1-357 #label HUL
#cross-references EMBL:X60295; NID:g21203; PID:g21204
GENETICS
```

```

#gene          secl
CLASSIFICATION #superfamily gliadin
FEATURE
1-19          #domain signal sequence #status predicted #label SIG\
20-357        #product omega secalin #status predicted #label MAT
SUMMARY       #length 357 #molecular-weight 41439 #checksum 3741

Query Match      2.3%; Score 229; DB 2; Length 357;
Best Local Similarity 28.2%; Pred. No. 8.19e-14;
Matches 71; Conservative 45; Mismatches 117; Indels 19; Gaps 17;

Db 15 IITAROLNPSGEELASPOOPVKEQ-S--YPOOPYPSHOPFTPOQSYSPYQPQFPQP 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1081 LVNGQALEPKQDAFGQEAAMVMDQKAGLYGQ-TYPAGQP-PMQGGF-HLQGSFSFNS 1137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 72 QQPAPIQOPFPFQOQPQFPFQOQQLPLQPQFPFQOPIPQOQSQFPPQOQRPQEQ 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1138 MM-NQMNOQGNFPLQGMHPRANTMRPTNTPKOLRMQLQRL--QGCG-FLNQS-RQALE 1192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 132 FPQPQOQIIPQIQOQFPFLQPOQGFPPQOQPSFAQPKQIISQOQPLQOQPFPSQPQP 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1193 LKWNPTAGGAANRPM-MQPOQGF-LNAQMV-AQKRELLSHH-FRQQRVAMMMQOQOQ 1248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 192 FPQPQGOIIPQOQPSPLQOQPFPSQOQRPQP-FPQO-QOQIIPQOQPFPLQPOQ 249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1249 -QQOQQOQ-QQQOQQOQQOQQOQTQAFSPPPNVTASPSMDGLLAGTWPQAPPQFPQPNY 1306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 250 PVQPQQRPFQ 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1307 GMSGQOQDPFAFR 1318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT      11
ENTRY
TITLE       S39356      #type complete
ORGANISM    transcription factor btd - fruit fly (Drosophila sp.)
DATE        #formal_name Drosophila sp.
            18-Feb-1994 #sequence_revision 01-Dec-1995 #text_change
            24-Sep-1998
ACCESSIONS S39356
REFERENCE    S39356
#authors    Wimmer, E.A.; Jaeckle, H.; Pfeifle, C.; Cohen, S.M.
#journal     Nature (1993) 366:690-694
#title       A Drosophila homologue of human Sp1 is a head-specific
            segmentation gene.
#accession   S39356
#status      preliminary
#molecule_type DNA
#residues    1-644  ##label WIM
#cross-references EMBL:229361; NID:g441283; PID:g441284
GENETICS
#gene        FlyBase:btd
#cross-references FlyBase:FBgn0000233
            245/2
#introns
SUMMARY      #length 644  #molecule-weight 68581  #checksum 6511

```

```

Query Match      2.3%  Score 230; DB 2; Length 644;
Best Local Similarity 48.3%  Pred. No. 6.01e-14;
Matches 43; Conservative 19; Mismatches 23; Indels 4; Gaps 4;

Db      31  QQQQHHLHMQQ-QAHHLHLSHQQAQQHMQHLTQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 89
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      1211  PQQGGFLNAQWVAQRSELLSHHFRCQVRVAMMQQQQQQQQQQQQQQQQQQQQQQQQQQQ-QQTTQ 1269

Db      90  HDFLSAALLSAPPSSLGSSGSSGSSSP 118

Qy      1270  -AF-SPPPNVTASPSMDGLLAGTVMQAP 1296

```

RESULT	12
ENTRY	S54522
TITLE	#type complete hypothetical protein YMR164c - yeast (Saccharomyces cerevisiae)
ALTERNATE NAMES	hypothetical protein YN8520.13c

[illegible]

```

RESULT      13
ENTRY
TITLE      A45266      #type complete
ENTRY      aryl hydrocarbon receptor - mouse
ALTERNATE_NAMES
ORGANISM    Ah receptor
#normal_name Mus musculus #common_name house mouse
DATE        21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
          20-Mar-1998
ACCESSIONS  A45266; JQ1485; A48601; A61274
REFERENCE
#authors    Burbach, K.M.; Poland, A.; Bradfield, C.A.
#journal     Proc. Natl. Acad. Sci. U.S.A. (1992) 89:8185-8189
#title       Cloning of the Ah-receptor cDNA reveals a distinctive
             ligand-activated transcription factor.
#cross-references MUID:92390411
#accession   A45266
             #status      preliminary
             #molecule_type mRNA: protein
             #residues     1-805 ##label BUR
             #cross-references GB:N94523; NID:q405913; PID:q192101
             ##note        sequence extracted from NCBI backbone (NCBIN:112849,
                             NCBIPI:112851)
REFERENCE    JQ1485
#authors     Ema, M.; Sogawa, K.; Watanabe, N.; Chujoh, Y.; Matsushita,
             N.; Gotoh, O.; Funae, Y.; Fujii-Kuriyama, Y.
#journal     Biochem. Biophys. Res. Commun. (1992) 184:246-253
#title       cDNA cloning and structure of mouse putative Ah receptor.
#cross-references MUID:92231934
#accession   JQ1485
             #molecule_type mRNA
             #residues     1-73, 'S', '75-131, 'FL', 134-170, 'HV', 173-805 ##label EMA
             #experimental_source HeLa-1 cells
             ##note        residues 10-25 have been confirmed by protein sequencing
REFERENCE    A48601
#authors     Schmidt, J.V.; Carver, L.A.; Bradfield, C.A.
#journal     J. Biol. Chem. (1993) 268:22203-22209
#title       Molecular characterization of the murine Ah gene.
             Organization, promoter analysis, and chromosomal
             assignment.
#cross-references MUID:94012821
#accession   A48601
             #status      preliminary; not compared with conceptual translation

```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 19:17:22 ; Search time 1111.79 Seconds
(without alignments)
7976.786 Million cell updates/sec

Title: US-09-041-994-1
Perfect score: 4496
Sequence: 1 GCTGGATGCTGACTACAGAG.....CATTTGAGCAGGAATCTTAG 4496

Scoring table: IDENTITY_NUC
2546578 seqs, 986266752 residues

Database : EST.*

- 1: em_est1.*
- 2: em_est2.*
- 3: em_est3.*
- 4: em_est4.*
- 5: em_est5.*
- 6: em_est6.*
- 7: em_est7.*
- 8: em_est8.*
- 9: em_est9.*
- 10: em_est10.*
- 11: em_est11.*
- 12: em_est12.*
- 13: em_est13.*
- 14: em_est14.*
- 15: em_est15.*
- 16: em_est16.*
- 17: em_est17.*
- 18: em_est18.*
- 19: em_est19.*
- 20: gb_est1.*
- 21: gb_est2.*
- 22: gb_est3.*
- 23: gb_est4.*
- 24: gb_est5.*
- 25: gb_est6.*
- 26: gb_est7.*
- 27: gb_est8.*
- 28: gb_est9.*
- 29: gb_est10.*
- 30: gb_est11.*
- 31: gb_est12.*
- 32: gb_est13.*
- 33: gb_est14.*
- 34: gb_est15.*
- 35: gb_est16.*
- 36: gb_est17.*
- 37: gb_est18.*
- 38: gb_est19.*
- 39: gb_est20.*
- 40: gb_est21.*
- 41: gb_est22.*
- 42: gb_est23.*
- 43: gb_est24.*
- 44: gb_est25.*
- 45: gb_est26.*
- 46: gb_est27.*
- 47: gb_est28.*
- 48: gb_est29.*
- 49: gb_est30.*
- 50: gb_est31.*
- 51: gb_est32.*
- 52: em_est20.*
- 53: em_est21.*

54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	534.4	11.9	547	34	AA488485	AA488485 ab39a08.r
2	385.2	8.6	437	26	W84822	W84822 zb53e05.r1
3	382.4	8.5	405	28	AA065270	AA065270 e08502r.T
4	370.0	8.2	425	28	AA114092	AA114092 zn66e10.r
5	353.8	7.9	427	29	AA150333	AA150333 z103e07.r
6	347.6	7.7	405	28	AA065268	AA065268 f05502m.T
7	343.4	7.6	457	40	AA920142	AA920142 vy52q12.r
8	316.4	7.0	405	28	AA065271	AA065271 f08502r.T
9	309.0	6.9	405	28	AA065272	AA065272 c08500r.T
10	300.0	6.7	464	21	T77368	T77368 yd72g08.r1
11	299.8	6.7	502	38	AA764263	AA764263 vv49f10.r
12	297.4	6.6	460	46	AI440499	AI440499 tc83e05.x
13	263.6	5.9	516	34	AA530243	AA530243 vj38c06.r
14	252.8	5.6	323	22	R28559	R28559 yb55b11.r1
15	247.8	5.5	287	27	AA045419	AA045419 zk59c01.r
16	245.0	5.4	449	26	W84775	W84775 zh53e05.s1
17	242.2	5.4	305	31	AA300819	AA300819 EST13767
18	229.2	5.1	405	28	AA065269	AA065269 d07502m.T
19	213.2	4.7	402	22	R21765	R21765 yb21a01.r1
20	193.6	4.3	377	22	R65357	R65357 y134g11.r1
21	180.2	4.0	436	41	AI044519	AI044519 UI-R-C1-k
22	167.8	3.7	176	32	AA360136	AA360136 EST69252
23	152.8	3.4	387	27	C03704	C03704 C03704 Huma
24	151.4	3.4	209	39	AA823647	AA823647 vr69b05.s
25	147.4	3.3	487	21	T77062	T77062 yb72g08.s1
26	146.2	3.3	353	22	H04364	H04364 y120e11.r1
27	131.6	2.9	357	22	R21718	R21718 yb21a01.s1
28	130.2	2.9	454	22	R28355	R28355 yb55b11.s1
29	128.0	2.8	392	22	R67499	R67499 y133g11.s1
30	128.0	2.8	392	40	AA946543	AA946543 EST202042
31	97.0	2.2	431	29	AA163058	AA163058 ms24d10.r
32	91.4	2.0	406	36	AA619146	AA619146 vc68e11.r
33	90.4	2.0	430	25	N64615	N64615 yz86e06.s1
34	90.0	2.0	368	49	AI664474	AI664474 ue63f06.r
35	87.4	1.9	524	38	AA740705	AA740705 ny98a04.s
36	83.2	1.9	385	22	R25318	R25318 yb42a03.r1
37	79.8	1.8	115	47	AI523909	AI523909 tg97h11.x
38	75.4	1.7	532	33	AA422600	AA422600 vd28e04.s
39	73.8	1.6	467	38	AA744255	AA744255 ny62g11.s
40	73.4	1.6	420	22	R66358	R66358 y134g11.s1
41	72.8	1.6	344	39	F23060	F23060 SSC19F03 Po
42	69.2	1.5	360	50	AI678308	AI678308 tu80e01.x
43	69.0	1.5	482	29	AA187197	AA187197 zp69h04.r
44	69.0	1.5	532	40	AA939852	AA939852 vz95f01.r
45	68.4	1.5	418	42	AI077892	AI077892 oy13a04.s

ALIGNMENTS

RESULT 1

AA488485

LOCUS

DEFINITION

ab39a08.r1 Stragene Hela cell s3 937216 Homo sapiens cDNA clone

IMAGE:843158 5' similar to TR:G1314285 GRIPI ; , mRNA

sequence.

ACCESSION

AA488485

NID

AA488485 547 bp mRNA EST 11-AUG-1997
ab39a08.r1 Stragene Hela cell s3 937216 Homo sapiens cDNA clone
IMAGE:843158 5' similar to TR:G1314285 GRIPI ; , mRNA
sequence.
ACCESSION
AA488485
NID
92215916

```

VERSION      AA488485.1  GI:2215916
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Primates; Catarrhini; Hominiidae; Homo.
              1 (bases 1 to 547)
              Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
              Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
              Moore, B., Scheinberg, K., Steptoe, M., Tan, F., Theising, B.,
              White, X., Wyllie, T., Waterston, R. and Wilson, R.
TITLE        WashU-Merck EST Project 1997
JOURNAL      Unpublished (1997)
COMMENT      On Jan 25, 1995 this sequence version replaced gi:637866.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m3 rev1 Et from Amersham
High quality sequence stop: 445.
Location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Stratagene HeLa cell s3 937216"
/sex="female"
/dev_stage="HeLa S3 cell line"
/lab_host="SOLR (kanamycin resistant)"
/notes="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dr. HeLa S3
epithelioid carcinoma cells grown to semi-confluency
without induction. Average insert size: 1.5 kb; Uni-ZAP XR
Vector. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3',
adaptor sequence: 5' CTCGATTTTTTTTTTTT 3'"

BASE COUNT  163 a 145 c 127 g 112 t
ORIGIN

Query Match      11.9%; Score 534.4; DB 34; Length 547;
Best Local Similarity 99.6%; Pred. No. 2.2e-135;
Matches 546; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

3046 TGAATCCCATGGGAATGGGGCTTAATCCCTATGGCCAGCAGCATCTAACCAACT 3105
1 TGAATCCCATGGGAATGGGGCTTAATCCCTATGGCCAGCAGCATCTAACCAACT 60

3106 GGGTTCTTGGCCGATGCATGTTCTTCATGGAACAAGTTCTCATGCACTCAAAATAG 3165
Db 61 GGGTTCTTGGCCGATGCATGTTCTTCATGGAACAAGTTCTCATGCACTCAAAATAG 120

3166 GCCTCTTTAGGAATTCCTGGATGATCTTGTGGGCCACCTTCCAACTCGGAAGGCCA 3225
Db 121 GCCTCTTTAGGAATTCCTGGATGATCTTGTGGGCCACCTTCCAACTCGGAAGGCCA 180

3226 GAGTCAGCAAGACGATATTGGACAGCTGCGACACTTTCTCAGCAACACAGATGCGAC 3285
Db 181 GAGTCAGCAAGACGATATTGGACAGCTGCGACACTTTCTCAGCAACACAGATGCGAC 240

3286 AGGCTCTGGAAGAAATTGACAGACTTTGGGCATTCCTGAACCTGTCATCGGGACAGC 3345
Db 241 AGGCTCTGGAAGAAATTGACAGACTTTGGGCATTCCTGAACCTGTCATCGGGACAGC 299

3346 ATTAGAGCCCAACAGGATGCTTTTCCAGAGCCCAAGCAGCAGTAATGATGGATCAAA 3405
Db 300 ATTAGAGCCCAACAGGATGCTTTTCCAGAGCCCAAGCAGCAGTAATGATGGATCAAA 359

3406 GGCAGGATTATATGACAGACATACCCAGCAGCGGGCTCCCAATGCAAGGAGGCTTTCA 3465

```

```

Db 360 GGCAGGATTATATGACAGACATACCCAGCAGGGGCTCCCAATGCAAGGAGGCTTTCA 419
Qy 3466 TCTTCAGGGACAATCACCATTCTTTTAACTCTATGATGAATCAGATGAACCAAGGCAA 3525
Db 420 TCTTCAGGGACAATCACCATTCTTTTAACTCTATGATGAATCAGATGAACCAAGGCAA 479
Qy 3526 TTTTCTCTCTCAAGGAATGCACCCAGCAGCAACATCATGAGCCCGGACAAACACCCC 3585
Db 480 TTTTCTCTCTCAAGGAATGCACCCAGCAGCAACATCATGAGCCCGGACAAACACCCC 539
Qy 3586 CAAGCAAC 3593
Db 540 CAAGCAAC 547

```

RESULT 2

```

W84822
LOCUS      W84822          437 bp  mRNA                      EST      27-JUN-1996
DEFINITION zh53e05.r1 Soares fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:415808 5', mRNA sequence.
ACCESSION  W84822
NID        91395952
VERSION    W84822.1  GI:1395952
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS     Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
            Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
            Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
            Trevasaki, E., Waterston, R., Williamson, A., Wohlmann, P. and
            Wilson, R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     On May 18, 1995 this sequence version replaced gi:811263.

```

```

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 312.
Location/Qualifiers
1..437
/organism="Homo sapiens"
/db_xref="GDB:1324278"
/db_xref="taxon:9606"
/clone="IMAGE:415808"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DHI0B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I oligo(dT) primer [5'
AATCGAAGAAATTAATAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

```

FEATURES

```

source
120 a 114 c 101 g 99 t 3 others
BASE COUNT
ORIGIN

```

Query Match 8.6%; Score 385.2; DB 26; Length 437;
 Best Local Similarity 96.6%; Pred. No. 1.3e-94;
 Matches 424; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

QY 3986 CCATATCAACCAAAATATGGAATGGGACAAACACAGATCCAGCTTTGGTCGAGTGTCT 4045
 Db 1 CCATATCAACCAAAATATGGAATGGGACAAACACAGATCCAGCTTTGGTCGAGTGTCT 60

QY 4046 AGTCCTCCCAATGAATGATGTCGTCAAGATGGTCCCTCCCAAGATCCCATGATGCAA 4105
 Db 61 AGTCCTCCCAATGAATGATGTCGTCAAGATGGTCCCTCCCAAGATCCCATGATGCAA 120

QY 4106 CACCCGAGGCTGATCATCTATCATCTCAGTATCAAGATGAAGGCTGCCATCAGGAAT 4165
 Db 121 CACCCGCA-GCTGATCATCTATCATCTCAGTATCAAGATGAAGGCTGCCATCAGGAAT 179

QY 4166 TTGGCCAGGAACAGCTCTCTTTCCACAGCAGCTTTCCACACAGGGGAATCCTGCAGTG 4225
 Db 180 TTGGCCAGGAACAGCTCTCTTTCCACAGCAGCTTTCCACACAGGGGAATCCTGCAGTG 239

QY 4226 TATAGTATGTCACATGAATGGCAGCAGTGGTCACATGGGACAGATGAACATGAACCCC 4285
 Db 240 TATAGTATGTCACATGAATGGCAGCAGTGGTCACATGGGA-AAATGAACATGAACCCC 298

QY 4286 ATGCCCATGTCGGCAGCTCTATGGGTCCTGATCAGAAATACCTGCTGACATCTCTGCACC 4345
 Db 299 ATGCCCATGTCGGCAGCTCTATGGGTCCTGATCAGAAATACCTGCTGACATCTCTGNACC 358

QY 4346 AGGACCTCTTAAGG--AAACACATGTAACAATGACATGACACTGACACTAGGATTTTGGGAAGGA 4403
 Db 359 AGGACCTCTTNAGGGAACACACTTGTNCAATGACACTGACACTAGGATTTTGGGAAGGG 418

QY 4404 ATCATTTGTCAGGCATCC 4422
 Db 419 ATCATTTGTCAGGCATCC 437

RESULT 3
 AA065270
 LOCUS e08502r Testis 5 Homo sapiens cDNA clone e08502 3' end, mRNA 25-SEP-1996
 DEFINITION sequence.
 ACCESSION AA065270
 NID g1929270
 VERSION AA065270.1 GI:1929270
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 405)
 AUTHORS Guellaen, G.
 TITLE Guellaen, G. Unpublished (1996)
 JOURNAL Unpublished (1996)
 COMMENT On Sep 1, 1995 this sequence version replaced.

CONTACT: Guellaen G
 Unite INSERM 99
 INSERM Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
 Tel: (33)149813530
 Fax: (33)148980908
 Email: guellaen@infobio.fr
 This sequence derives from a clone which was selected from the cDNA library - Testis 5 - using a repeat of 14 CAG as probe
 Seq primer: M13 reverse.
 Location/Qualifiers
 1..405
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="e08502"
 /clone_lib="Testis 5"

FEATURES
 source

/note="Vector: pSPORT1; Site_1: MluI; Site_2: NotI; mRNA was prepared from human testis of a 27 years old man. cDNA was prepared using a 15mer oligo dt anchored by two degenerated bases at its 3' end and containing a NotI site at its 5' end. The cDNA was cloned between SalI and NotI sites of pSPORT1. The MluI-SalI fragment came from the adaptor used for the cloning. The 3' end is at the NotI site. cDNA corresponding to abundant species were eliminated from this library."

BASE COUNT 112 a 101 c 103 g 87 t 2 others
 ORIGIN

Query Match 8.5%; Score 382.4; DB 28; Length 405;
 Best Local Similarity 97.8%; Pred. No. 7.3e-94;
 Matches 397; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 3044 GGTGAATCCCATGGGAATGGGGGCTAATCCCTATGGCCCAAGCAGCAGCATCTAACCAA 3103
 Db 1 GGTGAATCCCATGGGAATGGGGGCTAATCCCTATGGCCCAAGCAGCAGCATCTAACCAA 60

QY 3104 CTGGGTTCTTCGGCCGATGGCATGTTGTCCATGGAACAAGTTTCTCATGGCACTCAAAAT 3163
 Db 61 CTGGGTTCTTCGGCCGATGGCATGTTGTCCATGGAACAAGTTTCTCATGGCACTCAAAAT 120

QY 3164 AGGCCTCTTCTTAGGAATTCCTTGGATGATCTTTGGGCCACCTTCCAACTGGGAAGC 3223
 Db 121 AGGCCTCTTCTTAGGAATTCCTTGGATGATCTTTGGGCCACCTTCCAACTGGGAAGC 179

QY 3224 CAGAGTGACAAAGAGCATTTATGGACCACTGTCACACTTCTCAGCAACACAGATGGC 3283
 Db 180 CAGAGTGACAAAGAGCATTTATGGACCACTGTCACACTTCTCAGCAACACAGATGGC 239

QY 3284 ACAGGCTCTGAAGAAATTTGACAGAGCTTTGGGCATCTCTGAACCTTGTCAATCAGGACAG 3343
 Db 240 ACAGGCTCTGAAGAAATTTGACAGAGCTTTGGGCATCTCTGAACCTTGTCAATCAGGACAG 299

QY 3344 GCATTAGAGCCCAACAGGATGCTTTCCAGGCCAAGACAGCAGCATTAATGATGATCAG 3403
 Db 300 GCATTAGAGCCCAACAGGATGCTTTCCAGGCCAAGACAGCAGCATTAATGATGATCAG 359

QY 3404 AAGCGAGGATTTATGACAGACATACCCAGCAGAGGGGCTCCAA 3449
 Db 360 AAGCGAGGTTTATATGGACAGACTTACCCAGCAGAGGGGCTCCCA 405

RESULT 4

AA114092

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

On Sep 1, 1995 this sequence version replaced.

Contact: Wilson RK

QY 3540 GAATGCACCCACGA--GCCAATCATGTAGACCCCGGACAAACACCCCAAGCAACTTAG 3597
 |||||
 Db 300 GAATGCACCCACGAAGCCCAATCATGTAGACCCCGGACAAACACCCCAAGCAACTTAG 359
 |||||
 QY 3598 AAT-GCAGCTTCACGAGGCTGCAGGCCAGCAG 3631
 |||||
 Db 360 AATGCAGCTTCACGAGGCTGCAGGCCAGCAG 394
 |||||

RESULT 6
 AA065268/c 405 bp mRNA EST 25-SEP-1996
 LOCUS f05502m Testis 5 Homo sapiens cDNA clone f05502 3' end, mRNA
 DEFINITION sequence.
 ACCESSION AA065268
 NID g1929268
 VERSION AA065268.1 GI:1929268
 WORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 405)
 AUTHORS Guellaen,G.
 TITLE Guellaen,G. Unpublished (1996)
 JOURNAL Unpublished (1996)
 COMMENT On Sep 1, 1995 this sequence version replaced.

Contact: Guellaen G
 Unite INSERM 99
 INSERM
 Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
 Tel: (33)149813530
 Fax: (33)14980908
 Email: guellaen@infobiogen.fr
 This sequence derives from a clone which was selected from the cDNA
 library - Testis 5 - using a repeat of 14 CAG as probe
 Seq primer: -21 M13.

FEATURES
 source
 1. .405
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="f05502"
 /note="Vector: pSPORT1; Site.1: MluI; Site.2: NotI; mRNA
 was prepared from human testis of a 27 years old man. cDNA
 was prepared using a 15mer oligo dT anchored by two
 degenerated bases at its 3' end and containing a NotI site
 at its 5' end. The cDNA was cloned between SalI and NotI
 sites of pSPORT1. The MluI-SalI fragment come from the
 adaptor used for the cloning. The 3' end is at the NotI
 site. cDNA corresponding to abundant species were
 eliminated from this library."

BASE COUNT 98 a 100 c 98 g 102 t 7 others
 ORIGIN

Query Match 7.7%; Score 347.6; DB 28; Length 405;
 Best Local Similarity 97.08; Pred. No. 2.4e-84;
 Matches 350; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4131 AGTCCTCAGAAATGAAGGCTGGCCATCAGAAATTTGGCCAGGAACAGCTCTTTTCCC 4190
 |||||
 Db 405 AGTCCCCANAAATGAAGGCTGGCCATCAGGAATTTGGCCAGGAACAGCTCTTTTCCC 346
 |||||

QY 4191 AGCAGCAGTTGGCCACAGGGGAATCCTGCAGTGTATAGTATGGTCACATGAATGGCA 4250
 |||||
 Db 345 AGCAGCAGTTGGCCACAGGGGAATCCTGCAGTGTATAGTATGGTCACATGAATGGCA 286
 |||||

QY 4251 GCAGTGTACATGGGACAGATGAACATGAACCCATGCCATGCTGGATGCTATGG 4310
 |||||
 Db 285 GCNCTGTCTNCATGGGACAGATGAACATGAACCCATGCCATGCTGGATGCTATGG 226
 |||||

QY 4311 GTCTGTATCAGAATACTGTGTGACATCTCTGCACAGGACCTCTTAAAGAAACCACTGTA 4370
 |||||
 Db 225 GTCTGTATCAGAATACTGTGTGACATCTCTGCACAGGACCTCTTAAAGAAACCACTGTA 166
 |||||

QY 4371 CAAATGACACTGCACAGGATTTGGGAAGGAATCAATTTCCAGGCATCCATCTTGA 4430
 |||||
 Db 165 CAAATGACACTGCACAGGATTTGGGAAGGAATCAATTTCCAGGCATCCATCTTGA 106
 |||||

QY 4431 AGAAAGGACCAAGCTTTGAGCTCCATCAAGGCTATTTTAACTGATGTCATTTGAGCAGAA 4490
 |||||
 Db 105 AGAAAGGACCAAGCTTTGAGCTCCATCAAGGCTATTTTAACTGATGTCATTTGAGCAGAA 46
 |||||

QY 4491 T 4491
 Db 45 T 45

RESULT 7
 AA920142 457 bp mRNA EST 20-APR-1998
 LOCUS AA920142
 DEFINITION vY52q12.r1 Strata gene mouse lung 937302 Mus musculus cDNA clone
 IMAGE:1299142 5' similar to TR:009000 O09000 P300/CBP/CO-INTEGRATOR
 PROTEIN.; mRNA sequence.
 ACCESSION AA920142
 NID g3066921
 VERSION AA920142.1 GI:3066921
 WORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 457)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1407511.

Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LML; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:680190
 Possible reversed clone: similarity on wrong strand
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 410.

FEATURES
 Location/Qualifiers
 1. .457
 /organism="Mus musculus"
 /strain="C57BL/6 x CBA"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:1299142"
 /clone_lib="Stratagene mouse lung 937302"
 /sex="female"
 /tissue_type="lung"
 /dev_stage="6-8 month old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: lung; Vector: pBluescript SK-; Site.1:
 ECORI; Site.2: XhoI; Cloned unidirectionally. Primer:
 Oligo dT. 6-8 month old female lung and 1.5 year old male
 lung were source of mRNA. Average insert size: 1.5 kb;
 Uni-ZAP XR vector; -5' adaptor sequence: 5' GAATTCGGCAGAG
 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

BASE COUNT 120 a 119 c 122 g 96 t

ORIGIN

Query Match 7.6%; Score 343.4; DB 40; Length 457;
Best Local Similarity 84.5%; Pred. No. 3.6e-83;
Matches 386; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 3058 GGGATGGGGGCTAATCCCTATGGCAACAGCAGCATCTAACCACTGGTTCCTGGCC 3117
DB 1 GGGATGGAGTCAATCCCTATAGCCACAGTGGCTTAACCAACAGGTTCTCTGGCC 60

QY 3118 CGATGCGATGTTGTCATGGAACAAGTTCCTCATGSCACTCAAAATAGGCCCTCTCTTAG 3177
DB 61 AGAGGCGATGCTCTATGGAACAAGTTCCTCACGGTCTCAAAATAGGCCCTCTCTTAG 120

QY 3178 GAATTCCTGGATGATCTTTGGGCCACCTTCCAACTCTGGAAGGCCAGAGTGAGCAAG 3237
DB 121 AAATCTCTCTGGATGATCTCTTTGGGCCACCTTCTAACGCGAGAGGGCCAGAGTGAGGAG 180

3238 AGCATTTATTTGGACCGCTGCACACTCTTCTCAGCAACACAGATGCGACAGGCTTGSAGA 3297
DB 181 AGCTCTCTGGACCGCTGCACACTCTCAGCAACACAGATGCGACAGGCTTGSAGA 240

QY 3298 AATTGACAGAGCTTTGGGCACTTCTGAACTTGTCAATCAGGACAGGCTTATAGGCCCAA 3357
DB 241 GATCGACAGGCGCTTTGGGAATTCCTGAGCTCGTGAATCAGGACAGGCTTTGGAGTCCA 300

QY 3358 ACAGGATGCTTTCCAAAGGCCAAGAGCAGCAGTATGATGATGATGATGATGATGATGATGAT 3417
DB 301 ACAGGATGCTTTCCAAAGGCCAAGAGCAGCAGTATGATGATGATGATGATGATGATGATGAT 360

QY 3418 TGGACAGACATACCCAGCAGCGGCTCCAAATGAAGGAGGCTTCACTCTTCAGGGACA 3477
DB 361 TGGACAGACATACCCAGCAGCGGCTCCAAATGAAGGAGGCTTCACTCTTCAGGGACA 420

QY 3478 ATCACCATTCTTAATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3514
DB 421 GTCAACATGCTTTAACTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 457

RESULT 8

AA065271
LOCUS f08502r Testis 5 Homo sapiens cDNA clone f08502 3' end, mRNA 25-SEP-1996
DEFINITION sequence.
ACCESSION AA065271
NID 91929271
VERSION AA065271.1 GI:1929271
WORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 405)
AUTHORS Guellaen,G.
TITLE Guellaen,G. Unpublished (1996)
JOURNAL Unpublished (1996)
COMMENT On Sep 1, 1995 this sequence version replaced.

Contact: Guellaen G

Unite INSERM 99

INSERM

Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France

Tel: (33)149813530

Fax: (33)149890908

Email: guellaen@infobiogen.fr

This sequence derives from a clone which was selected from the cDNA library - Testis 5 - using a repeat of 14 CAG as probe

Seq primer: M13 reverse.

Location/Qualifiers

1..405

/organism="Homo sapiens"

/db_xref="taxon:9606"

FEATURES

source

/clone="f08502"

/clone_lib="testis 5"

/note="Vector: pSPORT1; Site_1: MluI; Site_2: NotI; mRNA was prepared from human testis of a 27 years old man. cDNA was prepared using a 15mer oligo dt anchored by two degenerated bases at its 3'end and containing a NotI site at its 5'end. The cDNA was cloned between SalI and NotI sites of pSPORT1. The MluI-SalI fragment come from the adaptor used for the cloning. The 3' end is at the NotI site. cDNA corresponding to abundant species were eliminated from this library."

BASE COUNT 117 a 124 c 91 g 72 t 1 others

ORIGIN

Query Match 7.0%; Score 316.4; DB 28; Length 405;
Best Local Similarity 91.8%; Pred. No. 8.2e-76;
Matches 368; Conservative 0; Mismatches 27; Indels 6; Gaps 3;

QY 3796 GAGGTGGCTATGATGTCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCA 3855
DB 3 GATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 62

QY 3856 ACAGCAACAGCAACAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3915
DB 63 ACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 122

QY 3916 GACTGCTTCCCGCAGCAGTGGGCTTTTGGCAGGACCCACAAATGCCACAAGCTCTCC 3975
DB 123 GACTGCTTCCCGCAGCAGTGGGCTTTTGGCAGGACCCACAAATGCCACAAGCTCTCC 182

QY 3976 CCAACAGTTTCCATATCAACCAATTTATGGAATGGGCAACACAGCAGCAGCAGCAG 4035
DB 183 GCAACAGTTTCCATATCAACCAATTTATGGAATGGGCAACACAGCAGCAGCAGCAG 242

QY 4036 TCGAGTGTCTAGTCTCTCCCAATGCAATGATGTCGTCAGAAGTGGTCTCCAGAAATCC 4095
DB 243 TCGAGTGTCTAGTCTCTCCCAATGCAATGATGTCGTCAGAAGTGGTCTCCAGAAATCC 302

QY 4096 CATGATGCAACACCGCAGGCTGCATCCATC---TATCAGTCTCTCAGAAATGAAGGGCT- 4152
DB 303 CATGATGCAACACCGCAGGCTGCATCCATC---TATCAGTCTCTCAGAAATGAAGGGGTG 362

QY 4152 GGCCATCAGGAATTTGG--CCAGGAACAGCTCTCTTTTCC 4190
DB 363 GGCCATCAGGAATTTGGGGCCAGGGAACAGCTCTCTTTTCC 403

RESULT 9

AA065272
LOCUS AA065272 405 bp mRNA EST
DEFINITION c08500r Testis 5 Homo sapiens cDNA clone c08500 3' end, mRNA sequence.
ACCESSION AA065272
NID 91929272
VERSION AA065272.1 GI:1929272
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 405)
AUTHORS Guellaen,G.
TITLE Guellaen,G. Unpublished (1996)
JOURNAL Unpublished (1996)
COMMENT On Sep 1, 1995 this sequence version replaced.

Contact: Guellaen G

Unite INSERM 99

INSERM

Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France

Tel: (33)149813530

Fax: (33)149890908


```

l: 7400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2072768"
/clone_lib="NCI-CCAP_CLL1"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH108"
/notes="Vector: pT7D3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'.
TCTTCAATCTGAAGTGGCGGCCGATGCTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into

```

the Not I and Eco RI sites of the modified pMT3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo.

BASE COUNT
ORIGIN

129 a 107 c 100 g 124 t

Query Match 5.6%; Score 297.4; DB 46; Length 460;
Best Local Similarity 99.7%; Pred. No. 1.4e-70;
Matches 298; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4193 CAGCAGTTGGCCACAGGGAATCCTCAGCTATAGTATGTCACATGAATGGCAGC 4252
DB 460 CAGCAGTTGGCCACAGGGAATCCTCAGCTATAGTATGTCACATGAATGGCAGC 401
QY 4253 AGTGGTCACATGGGACAGATGAACATGAACCCCATGTCGTCATGTCGTCATGGGT 4312
DB 400 AGTGGTCACATGGGACAGATGAACATGAACCCCATGTCGTCATGTCGTCATGGGT 341
DB 4313 CCTGATCAGAATACTGCTGACATCTCTGCACAGGACCTCTTAAGGAACCATGTACA 4372
DB 340 CCTGATCAGAATACTGCTGACATCTCTGCACAGGACCTCTTAAGGAACCATGTACA 281
QY 4373 AATGACATGCTAGGATTTGGGAGGAGTATGTTCCAGGCATCCATCTTGGAG 4432
DB 280 AATGACATGCTAGGATTTGGGAGGAGTATGTTCCAGGCATCCATCTTGGAG 221
QY 4433 AAAGGACCATGTTGAGCTCCATCAAGGGTATTTAAGTATGATGTCATTTGAGCAGGAT 4491
DB 220 AAAGGACCATGTTGAGCTCCATCAAGGGTATTTAAGTATGATGTCATTTGAGCAGGACT 162

RESULT 13
AA530243

LOCUS AA530243 516 bp mRNA EST 22-JUL-1997
DEFINITION vj38c06.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
clone IMAGE:931306 5' similar to TR:G1336160 G1336160 STERIOD
RECEPTOR COACTIVATOR 1A. ; mRNA sequence.

ACCESSION

AA530243

NID 92272949

VERSION AA530243.1 GI:2272949

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 516)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

Waterston,R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

On Sep 12, 1996 this sequence version replaced gi:1393273.

CONTACT: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:536226

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 335.

Location/Qualifiers

1..516

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:931306"

FEATURES

source

/clone_lib="Stratagene mouse diaphragm (#937303)"
/tissue_type="diaphragm"
/dev_stage="adult"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: diaphragm; Vector: pBluescript SK-; Site:1:

EcoRI; Site:2: XhoI; Cloned unidirectionally from mRNA

prepared from diaphragm muscle. Primer: Oligo dt. Average

insert size: 1.5 kb. Uni-ZAP XR vector; -5' adaptor

sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTT 3'

BASE COUNT 121 a 150 c 141 g 104 t

ORIGIN

Query Match 5.9%; Score 263.6; DB 34; Length 516;
Best Local Similarity 80.6%; Pred. No. 2.6e-61;
Matches 308; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 3996 CAAATTTAGTAATGGGACACACACACATCGAGCCTTTGGTCGAGTGTCTAGTCTCTCCA 4055
DB 1 CAAATTTAGTAATGGGACACACACACATCGAGCCTTTGGTCGAGTGTCTAGTCTCTCCA 60
QY 4056 ATGCAATGATGTCGTCAAGATGGTCCCTCCCAAGATCCCATGATCAACACCCGAGG 4115
DB 61 GTCAATGATGTCGTCAAGATGGTCCCTCCCAAGATCCCATGATCAACACCCGAGG 120
QY 4116 CTGATCCATCTATCAGTCTCAGAAATGAAGGCTGGCCATCAGAAATTTGGCCAGGA 4175
DB 121 CCACACCATGATCAGCCTTCAGATATGAAGGGTGGCGTCAGGACACCTGGCCAGGA 180
QY 4176 ACAGCTCCTTTTCCAGCAGAGTTTCCCAAGGGAATCTCTGACATCTCTGACATCTCT 4235
DB 181 ATGGCTCTTCCCAAGCAGAGTTTCTCCCAAGGGAACCTTCGACATCTCTGACATCTCT 240
QY 4236 TGCATATGATGGCAGCAGTGGTCACATGGGACAGATGAACATCAACATGATGATGATG 4295
DB 241 TGCATATGATGGCAGCAGTGGTCACATGGGACAGATGAACATCAACATGATGATGATG 300
QY 4296 CTGGCATGCTATGCTGCTCAGAAATGCTCTGACATCTCTGACATCTCTGACATCTCT 4355
DB 301 CTGGCATGCTATGCTGCTCAGAAATGCTCTGACATCTCTGACATCTCTGACATCTCT 360
QY 4356 AAGGAACCATGCTTACAAATGA 4377
DB 361 GTACAGATGACATGTCACAGGA 382

RESULT 14

R28559

LOCUS

DEFINITION

R28559 323 bp mRNA EST 25-APR-1995

YH5511.r1 Soares placenta Nb2HP Homo sapiens cDNA clone

IMAGE:133629 5' similar to contains TAR1 repetitive element ; mRNA

sequence.

ACCESSION

R28559

NID 9784694

VERSION R28559.1 GI:784694

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 323)

Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,

Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,

Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,

Trevaskis,E., Waterston,R., Williamson,A., Woldmann,P. and

Wilson,R.

The WashU-Merck EST Project

Unpublished (1995)

TITLE

JOURNAL

COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 630
High quality sequence stops: 314

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 630 Std Error: 0.00

Seq primer: M13RPI

High quality sequence stop: 314.

Location/Qualifiers

FEATURES

source

1. .323
/organism="Homo sapiens"
/db_xref="GDB:539364"
/db_xref="taxon:9606"
/clone="IMAGE:133629"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
AAGTGAAGAATTCGCGCGCAGGAATTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 86 a 73 c 86 g 73 t 5 others
ORIGIN
Query Match 5.6%; Score 252.8; DB 22; Length 323;
Best Local Similarity 97.5%; Pred. No. 1.8e-58;
Matches 277; Conservative 0; Mismatches 4; Indels 3; Gaps 2;

QY 4137 CAGAAATGAAGGCTGCCATCAGGAAATTTGGCCAGGACAGCTCTTTCCACGACG 4196
|||||
Db 1 CAGAAATGAAGGCTGCCATCAGGAAATTTGGCCAGGACAGCTCTTTCCACGACG 60
|||||

QY 4197 AGTTGCCACACGAGGGAATCTCTGAGTGTATAGTGTGCACATGAATGGCAGCAGTG 4256
|||||
Db 61 AGTTGCCACACGAGGGAATCTCTGAGTGTATAGTGTGCACATGAATGGCAGCAGTG 120
|||||

QY 4257 GTCATATGGGACAGATGACATGAACCCCATGCCCATGTCTGGCATGCTATGGGTCCTG 4316
|||||
121 GTCATATGGGACAGATGACATGAACCCCATGCCCATGTCTGGCATGCTATGGGTCCTG 180
|||||

QY 4317 ATCAGAAATACTCTGACATCTCTGACAGGACCTCTTAAGGAACACCACTGTACAAA-T 4375
|||||
Db 181 ATCAGAAATACTCTGACATCTCTGACAGGACCTCTTAAGGAACACCACTGTACAAA-T 240
|||||

QY 4376 GACATGCTACTAGGATATTGGG--AAGGAATCATGTGTTCCAGG 4417
|||||
Db 241 GACATGCTACTAGGATATTGGGAGGAGGANTCATTTGTTCCAGG 284
|||||

RESULT 15
AA045419
LOCUS
DEFINITION 287 bp mRNA EST 10-MAY-1997
IMAGE:487104 5', mRNA sequence.
ACCESSION AA045419
NID 91523621
VERSION AA045419.1 GI:1523621
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 {bases 1 to 287}

AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.

TITLE

Generation and analysis of 280,000 human expressed sequence tags

JOURNAL

Genome Res. 6 (9), 807-828 (1996)

MEDLINE

97044478

COMMENT

On Jan 25, 1995 this sequence version replaced gi:637753.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1867 Std Error: 0.00

Seq primer: 28M13 rev2 from Amersham

High quality sequence stop: 232.

FEATURES

source

1. .287
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="GDB:3760818"
/db_xref="taxon:9606"
/clone="IMAGE:487104"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AAGTGAAGAATTCGCGCGCAGGAATTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 78 a 77 c 68 g 63 t 1 others
ORIGIN

Query Match 5.5%; Score 247.8; DB 27; Length 287;
Best Local Similarity 97.9%; Pred. No. 4e-57;
Matches 282; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

QY 3985 TCCATATCAACCAATTTATGGAAT-GGGACAACACCAAGATCCAGCCTTTGTCGAGTGT 4043
|||||
Db 1 TCCATATCAACCAATTTATGGAATCGGGACAACAACCAAGATCCAGCCTTTGTCGAGTGT 60
|||||

QY 4044 CTAGTCTCTCCCAATGCAATGATGTCGTCGAAGATGGGTCCCTCCAGAAATCCCATGATGC 4103
|||||
Db 61 CTAGTCTCTCCCAATGCAATGATGTCGTCGAAGATGGGTCCCTCCAGAAATCCCATGATGC 120
|||||

QY 4104 AACACCCGCGAGCTGCATCCATCTATCATGTCCTCAGAAATGAAGGGCTGGCCATCAGGAA 4163
|||||
Db 121 AACACCCGCGAGCTGCATCCATCTATCATGTCCTCAGAAATGAAGGGCTGG-CATCAGGAA 179
|||||

QY 4164 ATTTCGCCAGGAAACAGCTCTTTTCCAGCAGCAGTGTGCCACACGAGGGAATCCTGCAG 4223
|||||
Db 180 ATTTCGCCAGGAAACAGCTCTTTTCCAGCAGCAGTGTGCCACACGAGGGAATCCTGCAG 239
|||||

QY 4224 TGTATAGTATGTTGCACATGAA-TGSCACAGCTGGTGCATGGGACAG 4270
|||||
Db 240 TGTATAGTATGTTGCACATGAAATGGCAGCAGTGGTCACATGNGACAG 287
|||||

Search completed: September 18, 1999, 00:52:22

Job time: 20100 sec

This Page Blank (uspto)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)